Table of Contents

List of Figures

List of Tables

List of Appendices

Acknowledgement

First and foremost, praises and thanks to God for giving me strength and guide in the writing of this thesis.

I would like to express my special appreciation and thanks to my supervisor and mentor Professor Geoff Hide, Professor of Parasitology, School of Environment and Life Sciences, University of Salford, Manchester. It would have been quite impossible to carry out this work without his guidance and consistent support.

Many thanks to my collaborators at the Edinburgh University of Edinburgh (headed by Dr Kim Picozzi) and the Woodchester Park (headed by Professor Richard (Dez) Delahay). Without their contribution, it would have been impossible to complete this project.

I would like to express my sincere gratitude to my research team and colleagues (Andrew Whiteoak, Tanja Lepore, Peter Shum, Jaro, Gaber, Jess, Hattan, Mohammed, Abdullah, May, Ana, Paz, Laura, Maria, Judy, Ashraf, Mona) at the University of Salford Biomedical Research Institute, for their moral support and encouragement at various stages of this project.

I am intensely indebted to Denise Rennie, Associate Head Academic, School of Environment and Life Sciences for her mentorship and consistent encouragement towards career progression. Thanks to dearest Toni for her support towards proofreading my work and encouragement at various stages during write-up of this thesis.

I am also very grateful to the director of MARG Education Port Harcourt International Limited, Mr Godson Ideozu (brother), and Kurumiema Miabiye (President/CEO Southern Field Development Partners) for support towards securing the scholarship to fund this project. Brother, you are indeed a star!

I am indebted to the former president of the Federal Republic of Nigeria, His Excellency President Goodluck Ebele Jonathan, for making education for Nigerian youths a top priority in his administration. My gratitude to the Office of the Special Advisor to the President on Niger Delta (OSAPND) for providing me with funding for carrying out this project.

Finally, I am very grateful to my mother Dame Patience Ideozu and all my siblings (Harvey, Helen-May, Fortune, Maureen, Rosemary, Godson, Francis, Glory, Ireju, Donatus, Hope, Harrison and Kingsley) whose support and everyday prayers made it possible for me to complete my project and present this thesis. They have sacrificed a lot for me and I am extremely grateful.

Statement of Authorship

I hereby certify that I am the sole owner of this thesis. The work contained in this thesis has not been previously submitted for a degree or diploma at any other University or Institution. To the best of my knowledge, this thesis contains no material previously published or written by another author except where due reference has been made in accordance with standard referencing practices.

Signed……..

Date……

Abstract

Trypanosomes are a major cause of disease and death in both animal and human populations. Resistance or susceptibility to African trypanosomes has been associated with TLR9 gene. Hence, the primary objectives of this study were to apply molecular tools to investigate trypanosome infections in British badgers and Nigerian cattle, and to investigate variation in bovine TLR9 gene in relation to their trypanosome infection status. The ITS Nested PCR method was used to detect trypanosomes in the British badgers and Nigerian cattle. Two novel hemi-nested PCRs targeting two bovine TLR9 CpG Islands were developed to derive genetic and epigenetic DNA sequence data of 72 African bovine samples. Twenty-nine out of 82 badger samples amplified tested positive for trypanosomes (35.4% prevalence). Analysis of sequence data showed the badgers were infected with *Trypanosoma (Megatrypanum) pestanai* and as expected our phylogenetic analysis shows the badger trypanosome to cluster together with *T. pestanai* (100% bootstrap support) in the *Megatrypanum* clade. Ten out of 80 Southern Nigerian cattle were shown to be positive for trypanosomes resulting in a 12.5% prevalence rate. A total of 9 polymorphisms were found in targeted region of bovine TLR9 gene. The study showed no significant association between SNPs and trypanosomiasis ($p = <0.05$). Our results confirm the methylation of bovine TLR9 gene and identified CpG-SNPs (2256 and 2865) which removes a C-G site and perhaps could alter DNA methylation as potential epigenetic markers for bovine TLR9 gene. Also, it reports the significant correlation between CpG Island SNPs ($p = <0.05$, all cases), suggesting possession of one Island SNP is a predicting tool for possession of the others. Future work is targeted at publishing papers in peer reviewed journals based on results from these studies.

Chapter 1

Introduction

The 20th and early 21st centuries have been marked by improvement in health that accompanied better nutrition and food handling, better housing and housing regulations, stronger legislation, a more enlightened attitude towards social problems and development of antibiotics and vaccines that greatly reduced infectious diseases in civilised societies (Playfair and Bancroft 2004). However, there are growing concerns about the persistence, emergence or re-emergence of a host of other infectious disease threats both in the developing and developed nations and trypanosome infections figure prominently among these infectious diseases (Lizundia et al. 2011, Muhanguzi et al. 2014).

Trypanosome infections are caused by trypanosomes belonging to the genus *Trypanosoma* which are blood parasites of vertebrates with several species posing as agents of human and/or livestock diseases (Hamilton et al. 2004). For example, the dreadful human African trypanosomiasis (HAT) or sleeping sickness is caused by *Trypanosoma brucei* species (Barrett et al. 2003, Brun et al. 2010) and trypanosomiasis in Eurasian badgers (*Meles meles*) is caused by *Trypanosoma pestanai* (Lizundia et al. 2011, Macdonald and Newman 2002). Growing concerns over animal and/or human infection with *Trypanosoma* are due to the deleterious health impacts associated with the disease consistent with the resulting economic consequences (Kristjanson et al. 1999). This highlights the need to investigate these diseases and employ measures that would be effective in mitigating these consequences. Hutchinson et al. (2003) describes effective disease control and management to depend heavily upon knowledge of the epidemiology of the disease, which in turn relies upon methods that adopt the practice of screening of both animal and human populations.

It is important to note that animal and human trypanosomiasis are parasitic diseases and one problem in parasite epidemiology is the difficulty in identifying the causative agent, especially where the strains of the parasites involved are morphologically very similar but possess unique characteristics as shown in some species of trypanosomes (Hide 1998). Interestingly, such challenges have been subdued by recent advances in functional genomics and the subsequent utilisation of PCR-based diagnostic methods to differentiate between species/subspecies of *Trypanosoma* (Cox et al. 2005). Additionally, PCR-based approaches have also been applied to study mammalian toll-like receptors (TLRs) and how they are crucial for recognition of invading pathogens and stimulation of the innate immune responses (Seabury et al. 2007). However, not much research has been carried out on the role of bovine TLRs in response to pathogenic components (Cargill and Womack 2007).

1.1 Human African trypanosomiasis

Human African trypanosomiasis (HAT), or sleeping sickness is listed as one of the world's neglected tropical diseases (NTDs) caused by protozoan parasites (genus *Trypanosoma*) which are transmitted by the bite of tsetse fly (genus *Glossina*) (Kennedy, 2013). Although, several species of *Trypanosoma* affect humans and animals only two are responsible for human disease, *Trypanosoma brucei gambesiense and Trypanosoma brucei rhodesiense* responsible for HAT and *Trypanosoma cruzi* responsible for Chagas disease (the neglected American Trypanosomiasis) in the Americas (Barrett et al. 2003). HAT occurs in 36 sub-Saharan African countries and roughly about 60 million people are estimated to be at risk of acquiring the disease between latitudes 14^{0} and 19^{0} south throughout these countries at risk (Brun et al. 2010). Although, the World Health Organisation (WHO) reported in 2009 that the number of new cases of HAT has dropped below 10,000 (Fig. 1.1) for the first time in 50 years (Simarro et al. 2011), it is essential to also note that the disease occurs in rural

communities in sub-Saharan Africa that probably would lack an advanced diagnostic capacity for HAT and there are concerns reported cases might have underestimated the true prevalence (Wastling and Welburn 2011).

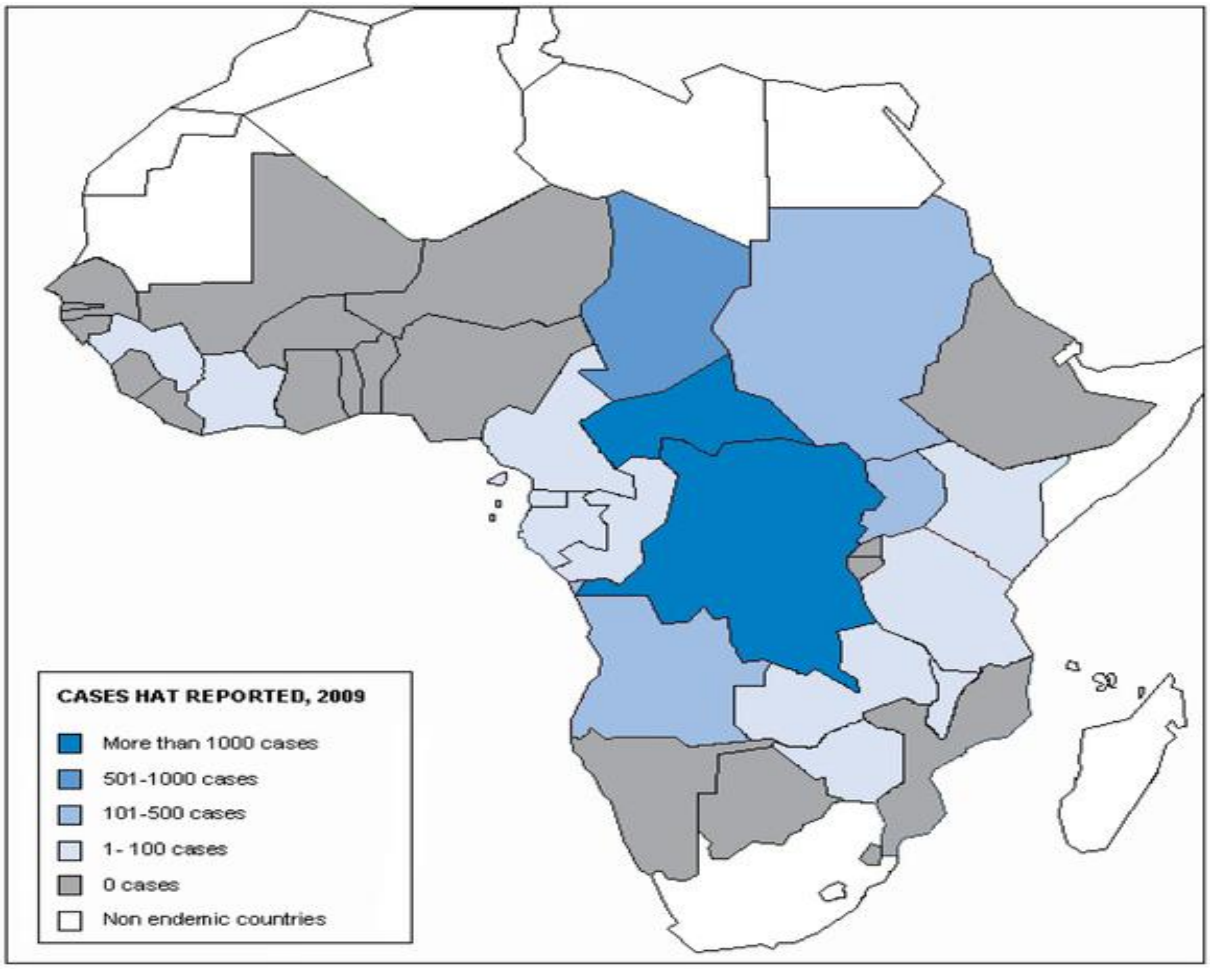


Figure 1.1 Categorisation of human African trypanosomiasis-endemic countries. Colour legends indicate cases of HAT in African countries according to WHO report in 2009 (Simarro et al., 2011).

Currently, the WHO targets the elimination of *T. b. gambiense* form of HAT as a public health problem in 2020 (WHO, 2014). However, it is worth noting that the disease is sporadic in nature due to the fact epidemics flare up and are interspersed with periods of endemicity (Hide 2003). An insight into three major epidemics of HAT that occurred which resulted in the death of hundreds of thousands human lives unveils this sporadic quality of the disease (Steverding 2008). More worrisome is the role of animal reservoir component for

transmission of HAT. For instance, the cattle have been shown to play host to the human infective *T. b. rhodesiense* parasite and this was implicated to be responsible for the 1984 Tororo District epidemic in Uganda, as it was shown isolates from sleeping sickness patients were similar and, in some cases, identical to those obtained from other regions 15 years earlier (Hide 2003).

Thus, while it is impressive and indeed a welcome development from the WHO report in 2009 confirming that the number of new cases of HAT has dropped below 10,000 for the first time in 50 years (Simarro et al. 2011), there is still evidence suggesting that some regions are experiencing a dramatic increase in the incidence rate of HAT. A recent study by Nmorsi et al. (2010), that investigated the incidence of HAT in Abraka communities within the Abraka Sleeping Sickness Focus (ASSF) in Nigeria confirmed that HAT infection in these communities were increasing and there was, in fact, ongoing transmission.

Hence, there is need to apply caution when interpreting the estimates of the true prevalence of the disease in Africa, rather more research should be encouraged to enhance enlightenment on factors that favour HAT epidemics such as the role of animal reservoir component. Nevertheless, the disease has only recently been implicated to be a major cause of rural underdevelopment in much of the sub-Saharan African countries where it persists and presents to be fatal if left untreated, with the severeness of the condition in an infected individual dependant on the form of HAT afflicted (Brun et al. 2010).

1.1.1 Symptoms of HAT

There are two forms of HAT caused by two human infective subspecies of *Trypanosoma brucei*, *Trypanosoma brucei rhodesiense* (*T. b. rhodesiense*) which is more common in East Africa and causes an acute form of the disease and *Trypanosoma brucei gambiense* (*T. b.*

gambiense) which is more common in West and Central Africa (Fig. 1.2) and is associated with long-term (chronic infection) human sleeping sickness (Hide 2003).

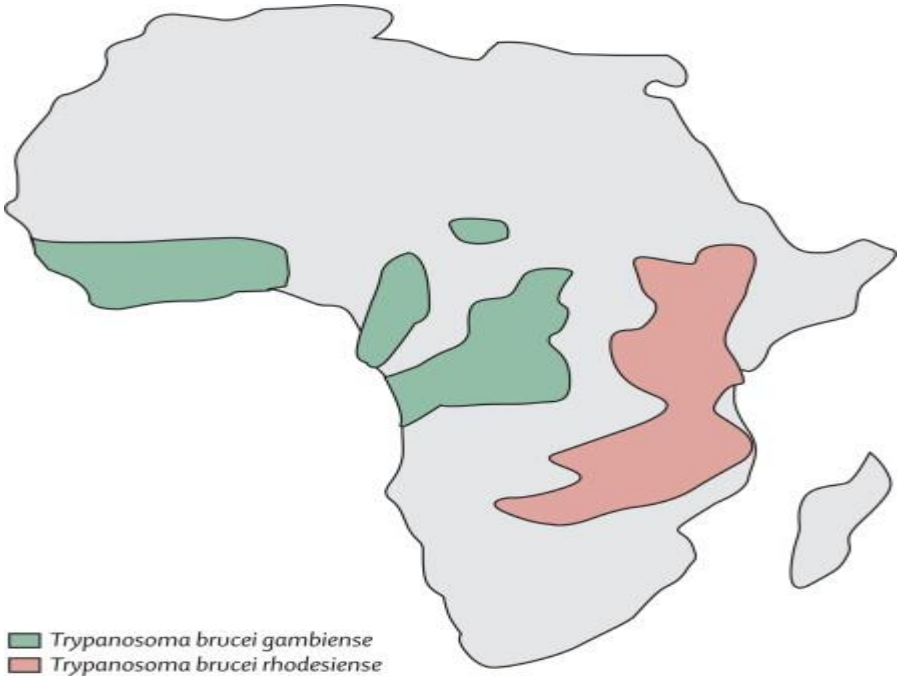


Figure 1.2 Representation of geographical distribution of *T. b. gambiense* and *T. b. rhodesiense* in Africa (Kennedy 2013).

Currently, *T. b. gambiense* is reported to account for 98% of HAT while *T. b. rhodesiense* accounts for only 2% of HAT (WHO, 2014). Traditionally, symptoms of HAT are described as occurring in early stages (stage 1) or late stages (stage 2) but it is difficult to distinguish both stages as they merge into each other (Kennedy 2013). However, following the bite of an infected tsetse fly the parasite multiplies in the lymph and the blood causing headaches, fever, fatigue, joint pains and stiffness, with the infected patient showing little or no signs of illness immediately (World Health Organization 2014). During the late stage of the disease, the infected patient suffers serious neurologic effects such as loss of coordination and disruption of biological rhythm which gives rise to inappropriate sleeping patterns (Barrett et al. 2003, Enanga et al. 2002). Hide (2003), distinguished between the two forms of HAT by identifying *T. b. rhodesiense* sleeping sickness to be fatal between 6-12 months while

infection with *T. b. gambiense* sleeping sickness could take up to 20 years before the patient finally succumbs.

1.1.2 Treatment of HAT

The type of treatment for HAT is dependent on the stage of the disease, as the first stage usually requires drugs of lower toxicity while the second stage requires more toxic drugs that can invade the blood-brain barrier to reach the parasite (World Health Organization 2014). WHO (2014) recommended drugs are presented in Table 1 below;

Table 1.1 WHO recommended drugs for the treatment of HAT

1ST Stage drugs	***T. b. gambiense* HAT**	***T. b. rhodesiense* HAT**
Pentamidine	Effective	Not effective
Suramin	Not effective	Effective
2nd Stage drugs		
Melarsoprol	Effective	Effective
Eflornithine	Effective	Not effective
Efflornithine+Nifurtimox	Effective	Not effective

Although, since 2009 nifurtimox-eflornithine combination therapy (NECT) has been adopted as the first line of treatment for 2nd stage *T. b. gambiense* HAT, two new drugs fexinidazole and SCYX-7158 are currently been evaluated in a clinical trial for treatment of 2nd stage *T. b. gambiense.* A successful development of these drugs may contribute towards achieving the goals of the recently launched WHO *T. b. gambiense* HAT elimination program (Eperon et al. 2014).

1.1.3 Life cycle of African trypanosomiasis

The blood-sucking tsetse fly is responsible for the transmission of both forms of HAT and infected domestic animals such as cattle act as a reservoir for the human infective *T. b.*

rhodesiense (Enyaru et al. 2006, Welburn et al. 2001). For *T. b. gambiense*, which is transmitted mostly from human to human (anthroponotic) (Brun et al. 2010), man act as the main reservoir but pigs and some other domestic animals have been reported as also being a reservoir (Njiokou et al. 2010). The tsetse fly takes in trypanosomes during a blood meal from an infected animal after which the parasites undergoes several morphological and biochemical changes in the fly's anterior midgut where the infection initially starts (Aksoy 2003).

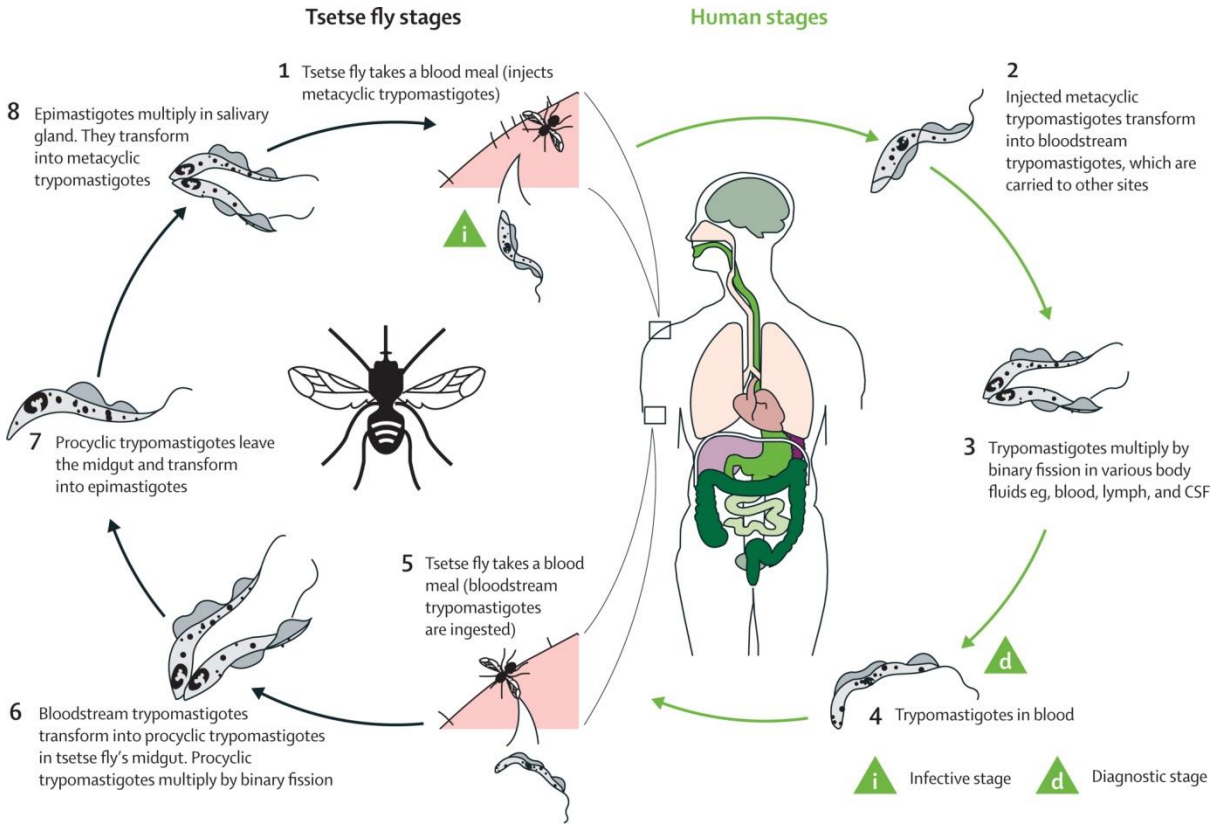


Figure 1.3 Diagrammatic representation of the life cycle of African trypanosomiasis (Kennedy 2013).

As shown in Fig. 1.3, the newly developed parasites in the tsetse fly's midgut then migrate to the salivary glands where they become epimastigotes, which then develops into short stumpy infective metacyclic trypanosomes which enter via the wound of a bitten individual, soon after which the trypanosome invades the bloodstream, lymph nodes and other tissues

(Kennedy 2004). The parasite is then transmitted to a subsequent host at the next blood meal (Brun et al. 2010, Hide 2003).

The life cycle of *Trypanosoma brucei* can be more complex, particularly due to the zoonotic nature of the disease as the two human infective subspecies can also be found across a range of domestic and wild animals (Brun et al. 2010). Additionally, there are sporadic reports emerging of disease in man caused by trypanosome species such as *T. b. brucei*, *T. congolense* and *T. evansi*, which normally are not pathogenic for human (Joshi et al. 2005, Truc et al. 1998). This gives a cause for concern as animal trypanosomiasis is not peculiar to Africa.

1.2 Animal trypanosomiasis: At a glance

Some species/subspecies of the *Trypanosoma* genes are pathogenic to domestic and wild animals in Africa resulting in animal African trypanosomiasis; they include *Trypanosoma brucei brucei, T. congolense, T. simae, T. suis and T. vivax* (Haore 1972). All these species result in diseases that are debilitating for animal health and are economically important, costing livestock producers and consumers a fortune to bear the burden. For instance, *T. b. brucei* is the causative parasite for animal trypanosomiasis in cattle (nagana) which in turn affects meat and milk production and economic losses due to the disease are estimated at $1340 million annually, excluding indirect livestock benefits such as manure and traction (Kristjanson et al. 1999). In a recent survey, trypanosomiasis was highlighted the most important problem for agriculture in Girji district of southern Ethiopia. Expenditure due to the disease was estimated for each household at $28 for preventive drugs and $18 for curative drugs (Chanie et al. 2013).

Nevertheless, some other species and subspecies of trypanosomes are prevalent elsewhere causing animal trypanosomiasis and continuously posing a major threat to animal and

possibly human health and the UK is no exception. Several species of trypanosomes have been reported in domesticated and free-living British mammalian fauna (Lizundia et al. 2011). For example, *T. theileri* affects British cattle and *T. melophagium* affects British sheep, with both belonging to the subgenus *Megatrypanum* and transmitted by tabanid flies and the sheep ked respectively (Bose and Heister 1993, Haore 1972). More worrisome is the evidence from a recent study by Lizundia et al. (2011), who detected that 31% of a sample of UK badger blood tested was infected with *T. pestanai* and identified the badger fleas (*Paraceras melis*) as the vector responsible for transmission. This gives a cause for concern as only recently there is a rising demand for culling of badgers in the UK due to its role as an animal reservoir for *Mycobacterium bovis* (*M. bovis*) which is the causative agent of bovine tuberculosis (bTB) (Corner et al. 2011). Additionally, with the rising sporadic reports of trypanosome infections in man caused by species of trypanosome that are pathogenic to animals and not humans (Joshi et al. 2005, Truc et al. 1998), there is need to investigate the UK badgers for presence of trypanosome and investigate what species/subspecies of *Trypanosoma* could be detected since it has only recently been highlighted by Lizundia et al. (2011) that research in this area is still sparse.

1.3 Animal trypanosomiasis in Nigeria

Nigeria which has gained recognition as Africa's most populous country ranked animal trypanosomiasis the fourth most important disease of cattle after rinderpest, contagious pleuropneumonia (CBPP) and dermatophilosis (Ademosun 1973, Bugaje 2006). Understandably, as cattle are a major livestock crucial for the supply of animal protein and any negative impact on their health would obstruct the supply of animal protein (Ogunsanmi et al. 2000).

However, the cattle health or existence is continuously threatened by trypanosomiasis with yearly losses in domestic animal production due to trypanosomiasis estimated at about N110m (currently $552,625.70) (Nigerian Institute for Trypanosomiasis Research 2008). Nigerian cattle have been reported to be infected with various species such as *T. b. brucei, T. vivax* and *T. congolense* (Enwezor et al. 2012, Omotainse et al. 2000) and the tsetse flies are the main vectors transmitting bovine trypanosomiasis in Nigeria, infesting as much as 80% of the nation's land mass (Abenga et al. 2004).

It has been reported that some of the factors that impact on the prevalence rate of trypanosomiasis in Nigeria are the species of tsetse fly involved in the transmission, the animal breed, type of management, season of the year and the type of vegetation (Ogunsanmi et al. 2000). For example, some riverine and forest species of the tsetse fly have been shown to be poorer vectors of trypanosome than their counterpart savannah species (Ikede et al. 1988). Although, cattle rearing occurs mostly in the northern part of the country, due to the nomadic practice involved in the Nigerian cattle market (Ogunsanmi et al. 2000) uninfected cattle may appear in regions infested with a high number of *Trypanosoma* infected tsetse that may predispose them to acquiring trypanosome infection or infected cattle may appear in regions infested with a high number of uninfected tsetse which may predispose them or even humans to trypanosome infection.

Finally, it is crucial to note that cattle movement was highlighted as the major facilitator of two epidemics of HAT that occurred in the Tororo District (1984) and Soroti District (1998) in south-east Uganda but thess were mainly due to *T. b. rhodesiense* (Welburn et al. 2001). In Nigeria where cattle are usually bred in the north but migrated to the south due to this nomadic practice (Ogunsanmi et al. 2000), *T. b. rhodesiense* HAT is not prevalent so cattle movement does not possess a great threat for HAT epidemic. Although, there has been a recent epidemic of HAT in the south, evidence from a recent research by Nmorsi et al. (2010)

that investigated the prevalence of HAT in the Abraka sleeping sickness focus (ASSF) of southern Nigeria confirmed the presence of *Trypanosoma* in some patients tested and stated there was, in fact, an ongoing transmission in the region but this was mainly caused by *T. b. gambiense*. They associated human activities such as farming and visiting of river as major risk factors for HAT infection. An earlier study by Airauhi et al. (2008) also confirmed the presence of *Trypanosoma* infection in this focus and addressed human hosts as the main animal reservoir.

Although, cattle movement does not facilitate the spread of human infective *T. b. gambiense* prevalent in Nigeria, the cattle in the northern part of Nigeria has been shown to play host to some other species of trypanosomes and this results in loss of animal production and a subsequent economic loss (Nigerian Institute for Trypanosomiasis Research 2008). Little is known on the prevalence of trypanosomiasis in the southern Nigerian cattle.

1.4 Diagnosis of trypanosomes

The basic requirement for any epidemiological study of a parasitic disease is the ability to identify the parasite or parasitic strain involved as it is vital for effective disease control and management, which in turn relies upon methods that incorporate screening of both animal and human populations for identification of disease-causing organisms (Hide 1998, Hutchinson et al. 2003). For instance, diagnosis of HAT depends heavily upon identification of the two species of *Trypanosoma* causing the disease, either *T. b. gambiense* or *T. b. rhodesiense* to avoid the risk of prescribing a more toxic drug to the less chronic form of the disease (Büscher 2014, Hide 2003).

The diagnosis of the two forms of HAT would rely on laboratory examinations based on the fact that the clinical features of the disease are not sufficiently specific (Chappuis et al. 2005). Clinical evidence of HAT infection relies on visualization of the parasite in blood, lymph or

cerebrospinal fluid (CSF) because HAT symptoms are initially not specific and the treatments are usually toxic (Wastling and Welburn 2011). A three stage approach which include screening, diagnostic confirmation and staging is used within control programmes and for individual patients (Brun et al. 2010).

The card agglutination test for trypanosomiasis (CATT) which was developed since in the late 1970s has been widely used for screening of human populations in Africa for infection with *T. b. gambiense* and can be done on serum, capillary blood obtained from a finger prick, as well as blood from impregnated filter papers (Magnus et al. 1978). CATT has been shown to be a fast and practical serological test which permits for large number of individuals to be screened daily for *T. b. gambiense* (Brun et al. 2010), however, it cannot be used for serological screening of *T. b. rhodesiense* and there are concerns over its limitations (Wastling and Welburn 2011).

Basically, CATT for *T. b. gambiense* is based on the variable surface antigen (VSG) type LiTat 1.3 (Magnus et al. 1978), but CATT false negatives have been observed in parasitologically positive cases of *T. b. gambiense* infection in Cameroon and were attributed to the absence of LiTat 1.3 antigen (Dukes et al. 1992). Also, CATT false negative was observed in some parasitologically positive cases in Uganda which tested positive by PCR for the gene encoding LiTat 1.3 (Enyaru et al. 1998). However, where there are good laboratory facilities ELISA, immunoflourescence assays and immune trypanolysis methods can be applied to detect anti-trypanosome antibody in sera (Chappuis et al. 2005).

Since screening methods are not 100% specific, it is essential for infection to be confirmed by detection of the parasite in the lymph of blood (Deborggraeve and Buscher 2010). During staging, progression of HAT is assessed by examination of CSF, for example, a patient is in stage one if the CSF white cell count is $\leq 5 \mu l^{-1}$ and there is no trypanosome observed while a

patient is in stage two when there is trypanosome and/or a white blood blood cell count of ≥ 20 μl^{-1} (Wastling and Welburn 2011).

1.5 Molecular diagnosis and phylogeny of trypanosomes

Molecular methodologies are becoming a vital tool for diagnosing human and animal trypanosomiasis due to their high specificity and sensitivity characteristics when detecting species/subspecies of *Trypanosoma* which was a major challenge previously associated with earlier diagnostic methods (Cox et al. 2005). Primers targeting various genes or proteins associated with trypanosomes are the focus of most molecular assays for diagnosing trypanosomiasis.

For example, the serum-resistance-associated (SRA) gene evolved by *T. b. rhodesiense* which makes it resistant to TLF-1 lysis (Capewell et al. 2011) has been the target for PCR amplification to diagnose the human-infective *T. b. rhodesiense* (Picozzi et al. 2008, Welburn et al. 2001). The variant surface glycoprotein (VSG) found in the surface of trypanosomes are responsible for their resistant to elimination by trypanolytic factor (TLF) (Mansfield and Paulnock 2005) and they have also been the target of PCR assays for diagnosing trypanosomiasis (Sengupta et al. 2010).

Other studies have targeted kinetoplastids DNA (Li et al. 2007) and ribosomal DNA (Cox et al. 2005, Desquesnes et al. 2011, Taylor et al. 2008). The rDNA which codes for rRNA is the catalytic component of the ribosomes in eukaryotes. The rRNA consists of tandem repeats, each repeat contains the 28S (large subunit), the 18S (small subunit), the 5.8S gene, the two internal transcribed spacer region (ITS1 and ITS2) and a large intergenic transcribed spacer (Richard et al. 2008). The ITS region of rRNA has highly conserved sequences that may be species-specific (Taylor et al. 2008), based on this they are a major target for molecular diagnosis of species/subspecies of trypanosomes using PCR. One technique that is based on

the variable ITS region of trypanosomes rRNA that has proved success in recent years is the ITS nested PCR. Cox et al. (2005) developed a nested PCR which amplifies the variable ITS region of the ribosomal gene locus using primers designed from a collection of 16 trypanosome species DNA sequences. The technique produces a unique size of PCR product for each species of trypanosome and it is capable of identifying mixed-species infection (Fig. 1.4).

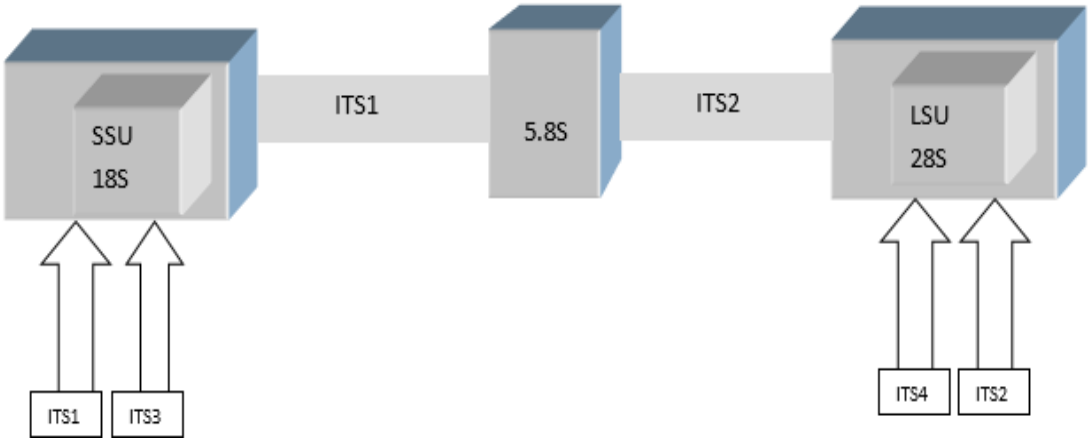


Figure 1.4 Regions of the ribosomal RNA gene locus targeted for nested PCR. The large boxes, small subunit (SSU) and large subunit (LSU) represent the conserved regions. The two spacers (ITS1 and ITS2) vary in size between species and occasionally subspecies. The nested primers are represented by black arrows (outer primers) ITS1 and ITS2 and white arrows (inner primers) ITS3 and ITS4 (Cox et al. 2005).

Nonetheless, there are other molecular methods that can be used to detect trypanosomes and these include mobile genetic elements PCR (MGE-PCR), loop-mediated isothermal amplification (LAMP) and nucleic acid sequence-based amplification (NASBA) (Hide and Tait 2009, Wastling and Welburn 2011). One major strength of the ITS nested PCR over other molecular methods is its capability to detect even mixed infections. It is worth noting that as there are a variety of trypanosome species causing diseases in different mammalian hosts (Table 1.2), the need for accurate identification of the species causing the infection is important for application of effective disease control measures (Hide 2003).

Table 1.2 Trypanosome species and hosts (Hoare, 1972).

Subgenus	Species	Hosts
Megatrypanum	*T. pestanai*	badgers
	T. melophagium	Sheep
	T. theileri	cattle
Herpetosoma	*T. otospermophili*	Ground squirrels
	T. lewisi	Black and brown rats
	T. microti	Field voles
	T. grosi	Wood-mouse
	T. musculi	House-mouse
	T. rangeli?	Human, mouse, rats
Incertae sedis	*T. petrodromi*	Elephant shrew
Pycnomonas	*T. suis*	Domestic and bush pigs
Schizotrypanum	*T. cruzi*	Human, monkeys, bugs
	T. dionisii/vespertilionis	Bats
Duttonella	*T. vivax*	Cattle, camels, goats, sheep
Nannomonas	*T. congolense*	Cattle, sheep, horses
	T. simiae	Pigs, cattle, camels
Trypanozoon	*T. brucei brucei*	Cattle, sheep, goats
	T. b. gambiense	Human
	T. b. rhodesiense	human
	T. evansi	Cattle, Indian elephants, dogs

As shown in Table 1.1 above, *T. rangeli* was classed as belonging to the *Herpetosoma* subgenus. Haore (1972) considered *T. rangeli* a distinct species of the subgenus *Herpetosoma*, some trypanosomes species belonging to this subgenus includes *T. microti*, *T. musculi*, *T. lewisi* and *T. grosi* (Sato et al. 2007). However, based on recent advances in phylogenetic analysis, *T. rangeli* has been shown to belong to a different clade and is more closely related to the *Schizotrypanum* species which includes *T. cruzi* and *T. dionisii* than to all other trypanosome species (Maia da Silva et al. 2004). Several phylogenetic studies have been carried out to explore evolutionary relations between trypanosome species (Gu et al. 2010, Maia da Silva et al. 2004, Stevens et al. 1999). Nevertheless, in the absence of fossil records to analyse the evolutionary trends or evolutionary relationships in some organisms including protozoa, phylogenetic analysis has emerged a critical tool that utilises datasets for studying the evolutionary relationships among genes and organisms by elegantly displaying them in a phylogenetic tree (Lemey et al. 2009). Molecular phylogenetic analysis is crucial for exploring taxonomic status of different species.

1.6 Host genetic variables in the determination of susceptibility or resistance to trypanosome infections

The normal human plasma contains trypanosome lytic factor (TLF) which destroys trypanosomes pathogenic for animals but each of the human infective species (*T. b. gambiense* and *T. b. rhodesiense*) are resistant to it by different mechanisms (Vanhamme and Pays 2004), which explicates reasons some animal trypanosomes are non-pathogenic for man (Capewell et al. 2011). The TLF has two components, apolipoprotein- L1 and haptoglobin-related protein functioning as destroyers for trypanosomes pathogenic to animals and non-pathogenic to humans (Shiflett et al. 2007) and an absence or impairment in any of these components may result in a human getting trypanosome infection pathogenic for animals. The reported case of human infection with *T. evansi* has been associated with the absence of apolipoprotein- L1, which also provides humans with innate immunity against trypanosomes pathogenic for animals (Vanhollebeke et al. 2006).

Meritoriously, the trypanosome genome which has now been sequenced is described to contain about 9,000 genes, of which, 10 % encode the variant surface glycoprotein (VSG) which are found on the surface of trypanosomes (Kennedy 2013). African trypanosomes are capable of avoiding the innate immune elimination by TLF via switching their VSG coat during infection (Mansfield and Paulnock 2005). Basically, during infection the trypanosome VSG is recognised by the host's immune system which triggers the production of immunoglobulin M (IgM) and immunoglobulin G (IgG) antibodies that neutralise the invading trypanosomes and reduce the level of parasitaemia but a few VSG would change their surface coat to a new VSG type not affected by the antibodies and this would continue until the immune system finally succumbs (Brun et al. 2010). It is on this basis Hide (2003) concluded that the prospects of developing a vaccine looks very slim.

Additionally, the human TLFs can be grouped into two; TLF-1 which contains the apolipoprotein- L1 and haptoglobin-related protein complex with a yet unknown mechanism through which it kills trypanosomes (Kieft et al. 2010). Nonetheless, the human pathogen *T. b. rhodesiense* has been highlighted to evolve the serum resistance protein (SRA), encoded by the SRA gene, which binds to the TLF-1 and making it resistant to lysis (Picozzi et al. 2008, Welburn et al. 2001). *T. b. gambiense* does not possess the SRA gene but has become resistant to TLF-1 using other mechanisms that involve reduced expression of the haptoglobin-haemoglobin receptor gene (Capewell et al. 2011).

1.6.1 Trypanotolerance

The epistemology of trypanotolerance was based on the observation that some cattle breeds have the capacity to better control trypanosome infection, as they survive, stay alive and remain productive despite a still active trypanosome transmission in endemic areas (Murray et al., 1982). Trypanotolerance is shown in some bovine breeds (*Bos taurus*) such as longhorn (N'Dama) and shorthorn (Baoule) cattle, which were introduced to Africa from the near East around 5000 BC so have closely evolved with trypanosome over the years, unlike the *Bos indicus* breeds that made their way to Africa only about 700 AD ago (Courtin et al. 2008, Murray et al. 1982).

Comparison of infections with *T. congolense* under laboratory conditions between the known trypanotolerant N'Dama calves (*Bos taurus*) with the more susceptible Boran calves (*Bos indicus*), confirmed observations in the field that N'Dama had remained productive and continued to gain weight (Paling et al. 1991). The trypanotolerant N'Dama breeds are believed to have a natural resistance due to their capability to reduce parasitaemia and anaemia which are the main pathogenic effects of the disease (Courtin et al. 2008). This natural selection has been shown to involve two mechanisms, an innate mechanism that

controls parasite growth and another mechanism involving haemopoietic system that acts to reduce anaemia (Naessens et al. 2002).

Also, some humans were reported to have exhibited trypanotolerance in regions of Africa. Findings from a 15 years follow-up of patients with *T. b. gambiense* HAT in Ivory Coast revealed that a few of the individuals that refused treatment became asymptomatic with no parasite detected in their blood through microscopy and PCR, and some of the patients eventually became seronegative and progressively negative to trypanosome variable antigens (Jamonneau et al. 2012). Although, there is no convincing evidence suggesting self-cures or trypanotolerance in *T. b. rhodesiense* HAT which is a more acute form of HAT (Courtin et al. 2008).

Considering the growing evidence emerging of trypanotolerance in cattle and human, there is a need to examine properly the innate immune responses and their role towards trypanosome infection. It was once described that the mammalian toll-like receptors (TLRs) play a crucial role in recognition of invading pathogens and stimulation of innate immune response (Seabury et al. 2007), an understanding of bovine TLRs and their role towards responding to *Trypanosoma* infection might be useful in building knowledge on trypanotolerance.

1.6.2 Innate immunity and mammalian toll-like receptors: their role in infectious diseases

The innate immune system which is composed of macrophages, natural killer cells, neutrophils, mucosal epithelial cells and endothelial cells plays a key role in providing protection within the first few minutes to hours of an infectious challenge (Abreu and Arditi 2004). Key advances in the field of innate immunity were triggered by the discovery that Toll receptors in *Drosophila* were important for clearance of pathogens (Lemaitre et al. 1996) and

the cloning of its mammalian counterpart, toll-like receptor 4 (TLR4) (Medzhitov and Janeway 2000).

Mammalian TLRs recognise molecular signatures of invaders known as the pathogen-associated molecular patterns (PAMPs) that are usually not expressed by the host (Medzhitov and Janeway 2000). Ten mammalian TLRs (TLR1-TLR10) have been identified so far and the TLR family is characterised by the presence of an extracellular leucine-rich repeats (LRRs) involved in ligand recognition and an intracellular cytoplasmic Toll/interleukin-1 (Toll/IL-1) receptor (TIR) domain that is akin to that of the IL-1 receptor family (Akira 2003).

The innate immune response can be stimulated by a wide range of bacterial, fungal, viral and protozoal component including lipopolysaccharides (LPS), peptidoglycan, glycolipids, and flagella (Abreu and Arditi 2004), when this happens individual TLRs recognise the structural components of pathogens specific to them and initiate an inflammatory response (Akira 2003). For example, LPS is a component of the outer membranes of Gram-negative bacteria which plays a potent activator of macrophages and is the causative agent of endotoxin shock (Akira 2003). After LPS has been released into the bloodstream, TLR4 is usually responsible for recognition and signalling it. Evidence from previous research (Poltorak et al. 1998) showed that mutations in TLR4 resulted in LPS unresponsiveness in mice.

Furthermore, TLR2 signalling is responsible for the killing of intracellular *Mycobacterium tuberculosis* in macrophages and absence of TLR2 in mice resulted in increased susceptibility to *Mycobacterium tuberculosis* (Reiling et al. 2002). Interestingly, TLR2 has also been identified to recognise glycosylphosphatidylinositol anchors from *T. cruzi* (Campos et al. 2001), the causative parasite for American trypanosomiasis which is also known as Chagas disease (Barrett et al. 2003). With the exception of TLR10 for which there is not yet a

specific ligand associated with it, ligands have been associated with all other TLRs and these are specified elsewhere (Akira and Takeda 2004).

Nevertheless, all mammalian TLR families function through the same signalling molecules including myeloid differentiation primary-response protein 88 (MyD88), IL-1 receptor-associated kinase (IRAK), tumor necrosis factor (TNF) receptor-associated factor (TRAF) 6, mitogen-activated protein (MAP) kinases and nuclear factor (NF) $_{k}$B (Akira, 2003). The signalling pathways of these receptors are shown below (Fig. 1.5).

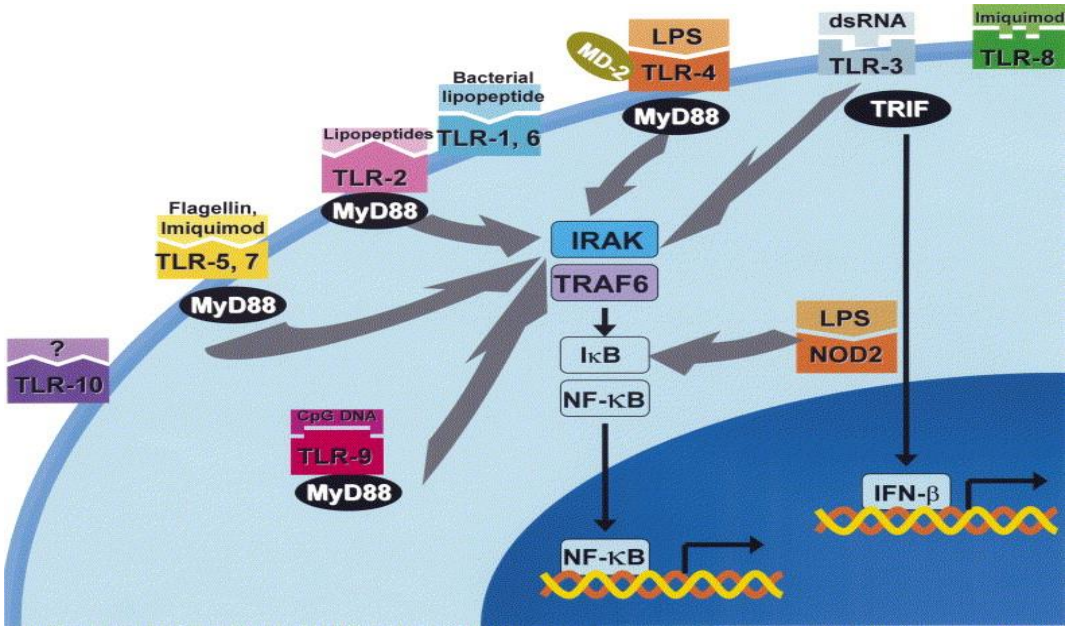


Figure 1.5 Representation of TLRs and their signalling pathway: Upon recognition of invading pathogenic components by TLRs, a signalling cascade is activated that employs several of the same molecules used by the IL-1 receptor. This induces several transcription factors which result in the activation of pro-inflammatory genes and a subsequent clearance of the pathogen (Abreu and Arditi 2004).

1.6.3 Bovine toll-like receptors

The advent of the Bovine Genome Project (https://www.hgsc.bcm.edu/content/bovine-genome-project) is an added prominence to the 21st century as it has contributed vastly towards research in understanding TLR genes in cattle. Using bovine sequence data from the

Bovine Genome Project has assisted researchers in establishing that all cattle 10 TLRs are homologous to those found in humans (Menzies and Ingham 2006). Radiation hybrid mapping is a genetic technique used in mapping mammalian chromosomes in the genome (Cox et al. 1990). Radiation hybrid mapping was used to map all 10 TLR genes in bovine genome; TLR1, TLR6 and TLR10 map closely together on *Bos taurus* chromosome 6 (BTA6), while TLR2, TLR3, TLR4, TLR5 and TLR9 map to BTA17, BTA27, BTA8, BTA16 and BTA22 respectively and TLR7 and TLR8 map to the X chromosome (McGuire et al. 2006).

Also, among cattle TLR3, TLR7, TLR8 and TLR9 genes, 139 polymorphisms have been identified, 130 of which are SNPs and the remaining 9 are insertion-deletion polymorphisms (INDELs) (Cargill and Womack 2007). Also, among cattle TLR1, TLR5 and TLR10 genes, 98 polymorphisms have been detected, 92 of which are SNPs and the remaining 6 INDELs, and 14 of all identified as non-synonymous SNPs positioned in domains of functional relevance (Seabury et al. 2007). Most recently 8 SNPs have been identified in cattle TLR2, TLR4 and TLR6 (Pariset et al. 2007).

Although, reports of bovine TLRs and their associations with pathogens were described as still scanty (Cargill and Womack 2007), evidence from a recent study showed that mastitis which is a disease caused by a bacterial infection resulted in an increased mRNA abundance of 3 genes viz. β–defensin 5, TLR2 and TLR4 but not TLR9 (Goldammer et al. 2004). However, TLR9 normally recognises pathogens of bacterial and viral infections (Akira and Takeda 2004). This implies that more research still needs to be done to investigate bovine TLRs and their associated ligands. Also, in another study Valarcher and Taylor (2007) implicated TLR3 and TLR4 signalling pathways to play a role in the activation of NF-$_K$B upon encounter with bovine syncytial viral infection. Lastly, a more recent study by Mucha et al. (2009) that investigated TLR gene mutation and increased susceptibility to

Mycobacterium paratuberculosis (MAP) infection in cattle, identified that two missense mutations in TLR4 (LRR domain) are associated with MAP infection. Interestingly, TLR9 has been associated with resistance to African trypanosomiasis in experimental mice (Drennan et al. 2005, Harris et al. 2007). However, despite all these efforts more research still needs to be done to build up a much needed evidence-based literature on bovine TLRs and how genetic or epigenetic alterations of these genes may influence resistance or susceptibility to infectious diseases, especially those specific for parasitic diseases as reports available in this area appear to be still scanty.

1.7 Epigenetics: The Role of DNA methylation in expression of TLRs

Epigenetics is a fast growing area of research which studies how environmental factors change the expression of genes without alteration in its inherited DNA sequence (Cai et al. 2015). Thus, studies in this area could provide new molecular insights for an indepth understanding of the epidemiology of infectious diseases. Epigenetic modifications of DNA and histones, the main components of chromatin, result in addition of extra information which influences the expression of the underlying genes (Law and Jacobsen 2010). DNA methylation, one such epigenetic modification, involves the covalent addition of methyl group via DNA methyltransferase (DNMT) enzymes to the cytosine residues of cytosine-phosphate-guanosine (CpG) dinucleotides in CpG Islands (Kulis and Esteller 2010). In eukaryotes, DNA methylation is the common epigenetic signalling tool that cells use to "switch off" genes crucial for several biological processes such as embryonic development, genomic imprinting, X-chromosome inactivation and preservation of chromosome stability (Phillips 2008). While in prokaryotes, DNA methylation provides a protection mechanism for host against digestion by restriction endonucleases designed to eliminate foreign DNA (Costello and Plass 2001).

DNA methylation occurs in about 70-80% of all CpG dinucleotides in mammalian genomes (Ehrlich et al. 1982), whereas the remaining unmethylated CG dinucleotides are found within promoters in CpG Islands (Illingworth and Bird 2009, Suzuki and Bird 2008). Methylation of CG sites in CpG Islands can lead to silencing of genes, a feature shown in cancer patients (Arai et al. 2006, Jones and Laird 1999). Also, a large number of tumor-specifically methylated genes were observed in patients with lung cancer and methylation of these genes was associated with loss of their protein expression (Heller et al. 2013). DNA methylation has also been reported to be involved in diseases such as Alzheimer's disease (De Jager et al. 2014), atherosclerosis (Lund et al. 2004), and type 1 diabetes (Stefan et al. 2014). Despite all the biological processes associated with DNA methylation, the exact role of DNA methylation in gene expression has not been well studied (Phillips 2008). While in mice deficiency in a particular DNA methyltransferase enzyme has been associated with decreased methylation levels and death during embryonic development (Suzuki and Bird 2008), some eukaryotes such as the yeast *Saccharomyces cerevisiae* and *Caenorhabditis elegans* are thought to have no methylated DNA at all (Phillips 2008). Nonetheless, little is known about the role of DNA methylation in the expression of TLR genes and studies in DNA methylation analysis with regards to infectious diseases are still scanty. Yet, there are several methods available to carry out such epigenetic studies (Hernandez et al. 2013, Shen and Waterland 2007). One study, however, reported higher methylation of TLR4 promoter in diabetic foot ulcer patients as compared to controls (Kanhaiya and Gupta 2013). Although, a recent study on genome-wide analysis of DNA methylation in bovine placentas identified the bovine TLR9 gene to be methylated (Su et al. 2014), the role of methylation in this gene is still not well understood. Also, research work on the impact of polymorphism in this methylated gene with regards to resistance or susceptibility of infectious diseases is still sparse.

1.8 Objectives

This project is aimed at using molecular tools for pan-*Trypanosoma* analysis and epigenetics of the host. To achieve this, molecular-based approaches using PCR methods will be developed and applied to investigate the prevalence of trypanosome infections in Eurasian badgers and Nigerian cattle, and to investigate genetic and epigenetic variation of bovine TLR9 gene in relation to their trypanosome infection status. The overall objectives of the project are grouped into three.

Firstly, the project will apply the ITS-Nested PCR to investigate the prevalence of trypanosome infections in Southern Nigerian cattle whilst simultaneously developing an agarose Gel-based Touch Preparation technique to enable a less laborious PCR re-amplification of any trypanosome strain initially detected. Two methods of DNA extraction from FTA cards (Single punches and multiple punches) would be compared to determine their effectiveness in estimating true prevalence (Chapter 3). Secondly, the project aims to investigate the prevalence of trypanosome infections in British badger blood samples from Woodchester Park in Gloucestershire using the ITS Nested-PCR. Where detected, an 18S rRNA PCR will be developed and applied to confirm the identity of species detected. Finally, to derive the 28S rRNA sequence from *T. pestanai* via PCR and apply molecular phylogenetic analyses using this marker to compare with the 18S phylogenies (Chapter 4).

Thirdly, the project is aimed at developing and applying molecular tools based on hemi-nested PCR to investigate the genetic and epigenetic variants in African bovine TLR9 genes for the identification of potential genetic and epigenetic markers which perhaps could be used to gain better understanding of susceptibility/resistance infectious diseases (Chapter 5). Reports from these studies may assist to build the literature in all areas investigated where there is an apparent gap in knowledge.

Chapter 2: Materials and methods

This research was targeted at the use of molecular tools to investigate trypanosome host-parasite interactions in mammalian populations. To achieve this, molecular-based approaches using PCR methods were developed and applied to investigate the prevalence of trypanosome infections in Eurasian badgers and Nigerian cattle, and to investigate genetic and epigenetic variation of bovine TLR9 gene in relation to their trypanosome infection status.

2.1 Sample characteristics

Firstly, to investigate the prevalence of trypanosomes in southern Nigerian cattle, a total of 80 bovine blood samples were obtained from bulls in an abattoir market in Ahoada, southern Nigeria. The blood from each sample was collected by butchers and placed on Whatman FTA cards (Chapter 3).

Secondly, to investigate the prevalence of trypanosomes in Eurasian badgers, blood was collected from 82 badgers in Woodchester Park, Gloucestershire, UK as part of a long-term study (Rogers et al. 1998). Data was collected on each badger and this included the sett where the badger was raised and where the badger now resides. For the purposes of analysis of evenness of trypanosome infection, the badger population was divided into North and South using a line of demarcation which was drawn separated by the Woodland Park ponds (Brick Kiln pond, Middle pond and Kennel pond) (See map, Fig. 4.2). The blood sample from each badger was stored at -20^{o}C (Chapter 4).

Lastly, to investigate the genetic and epigenetic variation in bovine TLR9 gene, a total of 72 cattle blood samples from northern Nigeria and Uganda were collected using Whatman FTA cards and provided by collaborators at Edinburgh University (Dr Kim Picozzi and colleagues). The Nigerian samples were taken from 3 villages (Tambes, Bokkos and Tambes)

while the Ugandan samples were from Abat village in Dakolo district of Uganda. These samples are of known trypanosome-infection status (*Trypanosoma* infected/ *Trypanosoma* non-infected) as tested by collaborators at the University of Edinburgh (Chapter 5).

2.2 Sample preparation for PCR

DNA from all badger blood samples were extracted using phenol-chloroform method while the bovine samples collected on FTA cards were either prepared using FTA purification reagent prior to direct PCR amplification or prepared using FTA purification reagent and then DNA eluted using Chelex resin extraction method prior to PCR amplification.

2.2.1 DNA Extraction: phenol-chloroform

Badger blood samples were extracted using a modification of a phenol-chloroform protocol as described previously (Morley et al. 2005, Morley et al. 2008) with appropriate measures to prevent contamination (Bajnok et al. 2014, Williams et al. 2005). 1ml of a coagulated badger blood was placed in a tube containing 400 µl lysis buffer and 100 µl proteinase K (20mg/ml) and incubated at 56^0C overnight. Then, 500 µl Tris-buffered phenol-chloroform (pH 8.0) was added to the tubes after incubation and mixed for 10 min before the two phases were separated by centrifugation for 10 min at 13,000rpm. The supernatant was transferred to a fresh tube and the step was repeated twice more to produce cleaner DNA. Then, to the supernatant 90ul of sodium acetate (3M pH5.3) and 900 µl of 100% ethanol was added before it was stored overnight at -20^0C. The solution was then centrifuged for 20 min at 13000rpm. The supernatant was discarded, washed in 500 µl 70% ethanol, centrifuged for another 10 min at 13,000rpm before the supernatant was discarded once more. The pellet on the bottom of the tube is the DNA which was dried at room temperature for 10 min before it was dissolved in 100 µl TE buffer, allowed to settle and become homogenous before storage

at 4^0 C or -20^0 C for long-term use. Using this protocol, DNA from all badger samples were successfully extracted and the results from the study presented in Chapter 4 of this report.

2.2.2 Preparation of FTA discs for PCR

All bovine blood used for this study were collected on Whatman FTA cards which must be prepared prior to PCR amplification. This is because the FTA card simplifies isolation, storage and processing of nucleic acids from various biological sources which include blood samples (http://www.fishersci.com/ecomm/servlet/fsproductdetail_10652_772329__-1_0). When the blood sample is applied to FTA card, it is allowed to dry and all cells are lysed then organisms deactivated. Then to prepare the FTA card for each of the bovine sample, the card was placed on a clean mat and a 2mm circular disc was taken by pressing Harris Micro Punch down and rotating according to manufacturer's instructions. This was placed in a sterile PCR amplification tube (0.5ml tube) and the punch cleaned by pressing several times on filter paper. Then, 200 µl of FTA purification reagent is added to the PCR amplification tube and the disc is giving a moderate wash by pressing the pipette up and down a couple of times before it is incubated at room temperature for 5 min. All used FTA purification reagent were removed and discarded using a pipette and this steps were repeated twice more for a total of 3 washes with FTA purification reagent. Following the above, 200 µl TE buffer (10 mM Tris, 0.1 mM EDTA, pH 8.0) was added to the tube, mixed and incubated for 5 min at room temperature. All used TE is removed and discarded and the steps repeated once for a total of 2 washes with TE buffer. The disc is then allowed to air dry for at least 1 hour before it is used to perform PCR amplification or stored at -20^0C for later use. Using this protocol, all southern Nigerian bovine samples were first prepared and a direct ITS-Nested was applied on them to investigate the prevalence of trypanosomes. The results are presented in Chapter 3 of this report. For the samples collected from Edinburgh, the discs were first prepared as above

but prior to PCR amplification the DNA was eluted using the Chelex extraction method described below.

2.2.3 Chelex Extraction from FTA cards

All bovine samples were collected on Whatman FTA cards. The discs from the cards were first prepared using the protocol described in Chapter 2.2.2 but prior to PCR amplification DNA was eluted using the Chelex extraction method described previously (Ahmed et al. 2011) but with a modification based on the use of single FTA card punches in this study. After the 2mm discs had been dried, the Chelex 100® elution resin was used to elute the DNA. 10% Chelex® solution (Sodium form, 100-200 mesh; Bio-Rad Laboratories, Hemel Hempstead, Hertfordshire, UK) was made by weighing out 5g of Chelex and adding it to a 50ml sterile falcon tube placed on a scale that has been balanced to zero. Then water was added up to the 50ml mark. This was then transferred to a sterile jar with a magnetic stirrer inside and heated on a hot plate for at least 15 minutes. One hundred microlitres (for multiple punches) or 20 µl (for a single punch) of 10% Chelex® solution was then added to the sample containing either multiple punches or single punch of FTA cards before it was mixed thoroughly by pipetting several times. Samples were heated for 30 minutes at 90˚C and stored at -20˚C until needed for further processing. Chelex extraction using multiple punches was applied to the Southern Nigerian bovine samples (Chapter 3) while Chelex extraction using single punches was applied to all bovine samples provided by collaborators at the University of Edinburgh (Chapter 5).

2.3 PCR-based DNA quality test

Extracted DNA was tested for the presence of and ability to amplify DNA using PCR and a set of mammalian tubulin primers that anneals to DNA of any mammal. This PCR has been

successfully tested on humans, rats, mice, sheep, cattle, voles, badgers, foxes, deer and bats. The following mammalian tubulin primers were used; Mammalian tubulin forward (5'- CGT GAG TCG ATC TCC ATC CAT-3') and mammalian tubulin reverse (5'- GCC CTC ACC CAC ATA CCA GTG-3'). PCR amplification was carried out using a Robocycler (Stratagene) and the reaction volume of 25 µl contained the following components; NH_4 buffer (Bioline, London) (final concentration of 16mM $(NH_4)\ 2SO_4$, 67 mM Tris-HCl, pH 9.0, 0.01% Tween -20), 5Mm $MgCl_2$, 5 µM of each of the primers (mammalian tubulin forward and mammalian tubulin reverse), 1 mM total dNTP's, 1.25 U of Biotaq (Bioline, London). The FTA punch that has been prepared or 1 µl of badger DNA was added to the reaction to make the final volume of 25 µl before the reaction was carried out in the following conditions: 5 min at 94^0C; 40 cycles x 40 s at 94^0C, 40 s at 60^0C, and 90 s 72^0C; 10 min at 72^0C. Following this gel electrophoresis was carried out and, where applicable, DNA sequencing and analysis was conducted. This technique was applied in Chapter 4 of this report to ensure newly extracted DNA from badger samples would amplify when used in PCR. It was also used to verify the quality of other DNA samples for troubleshooting as required.

2.4.1 ITS-nested PCR

Extracted DNA or FTA punches derived from this study are tested for trypanosome infection using the ITS-Specific (nested) PCR method with the following set of primers; the outer primer sequences were ITS 1 (5'-GAT TAC GTC CCT GCC ATT TG-3') and ITS 2 (5' TTG TTC GCT ATC GGT CTT CC-3') (MWG Biotech), and the inner primer sequences were ITS 3 (5' GGA AGC AAA AGT CGT AAC AAG G- 3') and ITS 4 (5' TGT TTT CTT TTC CTC CGC TG-3') (MWG Biotech) (Cox et al., 2005). The Robocycler (Stratagene) was used to perform the nested PCR amplification which usually involves 2 rounds of reactions and the

first round reaction volume of 25 µl consisted of the following components; KCl buffer (Bioline, London) (final concentration of 10 mM Tris- HCl, pH 9.0, 1.5 mM $MgCl_2$, 50 mM KCl, 0.1% Triton X -100, and 0.01% (w/v) stabilizer), 2 µM of each outer primer ITS1 and ITS2, 1 mM total dNTP's, 1.25 U of Biotaq (Bioline, London) and 5-15 ng/µl trypanosome infected genomic DNA or a prepared 2mm disc from FTA cards. The thermocycling conditions for the reaction were set as follows; 5 min at 94^0C; 35 cycles X 40 s at 94^0C, 40 s at 58^0C and 1 min 30s at 72^0C; 5 min at 72^0C.

For the second round reaction, 1 µl of the resulting PCR products from the first round reaction were placed in fresh tubes containing 24 µl of the reaction mixture (as detailed in the first round reaction mixture), with the exception of the substitution of the outer primers (ITS 1 and ITS 2) with the inner primers (ITS 3 and ITS 4). Following this, electrophoresis was commenced and where there is an indication for trypanosome infection in samples under UV transilluminator, the band size is used to estimate the species of trypanosomes detected (Cox et al. 2005). This technique was applied to detect trypanosomes in the badgers (Chapter 4) and the southern Nigerian cattle (chapter 3).

2.4.2 ITS5 and 6 PCR

The PCR method targets the internal transcribed spacer region of *Trypanosoma* rRNA. The PCR was designed as a third round PCR to amplify PCR products derived using the ITS-Nested PCR mentioned above (2.4.1). DNA sequences for various African trypanosomes (AF306771.1, AF306775.1, AF306773.1, AF306772.1, AF306774.1, AF306776.1, AF 306777.1, U 22315, U 22316, U 22317, U 22318, U 22319 and U 22320) were assembled from the national centre for biotechnology information (NCBI) database (http://www.ncbi.nlm.nih.gov/). The sequences were aligned with Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) before selected conserved regions were checked

for specificity using the Primer-BLAST tool on NCBI. The sequences were same used by Cox et al. (2005) for the design of ITS-Nested PCR. The forward primer sequence (ITS5) is; ITS 5: 5'-GTA GGT GAA CCT GCA GCT G- 3' while the reverse sequence (ITS6) is; 5'-GCT TAA GTT TAG CGG GTG GTC- 3'. The Robocycler (Stratagene) was also used to perform the ITS5 and 6 PCR amplification and reaction conditions and mixture stays the same as the ITS-Nested PCR. This method was applied in Chapter 4 to confirm another set of primers was also detecting trypanosomes from badgers.

2.5 18S PCR

Primers targeting the 18S region of trypanosomes were designed by assembling the 18S sequences of some *Herpetosoma* trypanosome species including *T. otospermophili* and *Megatrypanum* species including *T. pestanai*, accessible from the NCBI database. Sequences were aligned with Clustal Omega and selected conserved regions were checked for specificity using the Primer-BLAST tool on NCBI. The forward primer (18sF) is 5'-CAT GCA TGC CTC AGA ATC ACT G- 3' while the reverse primer (18sR) is 5'-CTG TTG CCC AAA ATC TCA CCT TGC-3'. The 24 µl reaction mixture for the PCR consists of MyTaq^{TM} HS Mix (Bioline, London) and 5 µM of forward and reverse primers. The amplifications were carried out using 1 minute at 95^{0}C, followed by 35 cycles of 15s at 95^{0}C, 15s at 56^{0}C, 1.30 minutes at 72^{0}C and a final cycle of 5 min at 72^{0}C. The technique was applied to badger samples for sequencing purposes to identify the species of trypanosomes detected in the badgers (Chapter 4).

2.6 28S PCR

The technique was developed for this project. The 28S ribosomal RNA of trypanosomes was the target region for PCR amplification. The 28S sequence for *T. pestanai* was not available

in the databases. Suitable primers for novel amplification of the 28S rRNA gene from *T. pestanai* were designed by assembling 28S sequences of trypanosomes; *T. otospermophili* (GI: 46091661), *T. kuseli* (GI: 46091662), *T. rangeli* (GI: 662247341) and *T. minasense* (GI: 159157536), accessible from NCBI and following the procedure described above. The forward primer (28S1F) is 5'-CAGACCTGAGTGTGGCAGG-3' while the reverse primer (28S1R) is 5'-AAGAAAGCTCACCGTGGGAGG-3'. The 24 µl reaction mixture for the PCR consisted of MyTaq^{TM} HS Mix (Bioline, London) and 5 µM of forward and reverse primers. The reaction was carried out in the following thermocycling conditions of 1 min at 95^{0}C, followed by 25 cycles of 15s at 95^{0}C, 15s at 54^{0}C, 1.30 min at 72^{0}C and a final cycle of 5 min at 72^{0}C. The technique was applied in chapter 4 to derive a novel 28S gene for species of trypanosomes detected in badgers (see chapter 4).

2.7 TLR9 coding region PCR

The primary target was to investigate sequence variation in the coding region which is within the mRNA of the TLR9 genome. A multiple alignment was performed on Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) using DNA sequences of 9 bovine TLR9 genomes; 7 Bos Taurus (EF076731.1, EF076724.1, EF076725.1, EF076727.1, EF076728.1, EF076729.1 and EF076723.1) and 2 Bos indicus (EF076726.1 and EF076730.1) sequences assembled from NCBI (http://www.ncbi.nlm.nih.gov/). This confirmed that except for minor mutations the sequences were homologous (Appendix G). Using the NCBI database the positions of mRNA and coding region were identified so primers were picked to cover these regions. The primers were designed acknowledging primer design tips from Life Technologies (http://www.lifetechnologies.com) and the specificity of primers were checked on Primer Blast (http://www.ncbi.nlm.nih.gov) before they were ordered from Eurofins MWG Operon (http://www.eurofinsgenomics.eu/).

The TLR9 forward primers sequence is; 5' – GGA GAA GCC GCA TTC CCT G - 3' while the reverse primer sequence is; TGT GGG GTT AAA GGA GTG CTG- 3' with both 5' for each primer located at position 606 and 4761bp respectively (Accession no: EE076731). The PCR amplification is carried out using the Robocycler (Stratagene). The 24 µl reaction mixture consists of Myfi ™ mix (ready-to-use 2x mix; Bioline, London) and 5 µM of forward and reverse primers. The reaction was carried out in the following thermocycling conditions of 5 min at 95^0C, followed by 35 cycles of 30s at 95^0C, 45s at 55^0C, 1.30 min at 72^0C and a final cycle of 5 min at 72^0C. Following this, electrophoresis is commenced and upon confirmation of bands of interest under UV transilluminator, PCR product is sent to GATC Biotech (http://www.gatc-biotech.com) for purification and sequencing. The TLR9 PCR was also applied in Chapter 3 to confirm if DNA samples collected on FTA cards were of bovine origin.

2.8 TLR9 Hemi-nested PCR

The primary target for the development and application of the hemi-nested PCR in this study was to amplify regions of the bovine TLR9 gene covering the two CpG Islands. The hemi-nested PCR is a double round PCR that first uses the TLR9 Coding Region PCR primers; TLR9F (5' – GGA GAA GCC GCA TTC CCT G - 3') and TLR9R (5'- TGT GGG GTT AAA GGA GTG CTG- 3') in a first round PCR for the amplification of the bovine TLR9 coding region using the same reaction conditions described for TLR9 coding region PCR above. From which, 1ul of the PCR product is used for a second round PCR using a new internal forward primer; TLR9 CpGF (5'–CTGCGTCTCCGGGACAATAAC-3') and the same internal primer TLR9R (5'- TGT GGG GTT AAA GGA GTG CTG- 3'). The PCR amplification is carried out using the Robocycler (Stratagene). The technique is, therefore, more sensitive and specific and is capable of producing more quality amplicons especially

where the starting material is low in concentration. The 24 µl reaction mixture consists of MyTaq™ Hs ready-to-use reaction mixture (Bioline, London) and 5 µM of forward and reverse primers. The reaction was carried out in the following thermocycling conditions of 10 min at 94^0C, followed by 35 cycles of 30s at 95^0C, 30s at 60^0C, 30s at 72^0C and a final cycle of 7 min at 72^0C. Following this, electrophoresis is commenced and upon confirmation of successful amplification by presence of bands after view under UV transilluminator, PCR product is sent to GATC Biotech (http://www.gatc-biotech.com) for purification and sequencing.

2.9 Bisulfite Conversion

A common method of detecting/quantifying DNA methylation is using bisulfite treatment of which DNA which converts unmethylated cytosines into uracils. Methylated cytosines remain unchanged during treatment, so after conversion the methylation status of DNA can be confirmed by bisulfite PCR and DNA sequencing (http://www.zymoresearch.com). The EZ DNA Methylation-Gold™ Kit (Zymo Research, Orange, CA, USA) was used to convert bovine DNA to bisulfite treated DNA. The bovine DNAs used were eluted using the Chelex extraction from FTA cards using the method described above. The DNA for treatment was made to 20 µl according to manufacturer's recommendation by adding 5 µl of Chelex eluted DNA from FTA card to 15 µl distilled water to make up a final volume of 20 µl in a PCR tube. Then 130 µl of CT Conversion Reagent was added to the tube which was placed in a thermal cycler (Prime Elite, Techne, UK) using a modified reaction condition at 95°C for 10 min and at 64°C for 2.5 hours described by Su et al. (2014). Other steps were followed as described in the manufacturer's instructions and the DNA was eluted according to the manufacturer's recommendation.

2.10 TLR9 bisulfite hemi-nested PCR

The TLR9 bisulfite PCR uses a hemi-nested approach to amplify the two bovine TLR9 CpG Islands which were the main targets for detecting and analysing DNA methylation in bovine TLR9 gene in this study (Chapter 5). Two sets of primers each was selected for both CpG Island 1 and CpG Island 2 using Methprimer (www.urogene.org/methprimer). For CpG Island 1, the forward primer is bCpGF1 (5'-GGAAATTAGTTGAAGGTTTTGAGTAA-3') while the reverse primer is bCpGR1 (5'-CTATCTCTCCAACAAAAAATCCAC-3'), then for CpG Island 2 the forward primer is bCpGF2 (5'-TTGGGATTTTTGGTATTGTTTTTAT-3') while the reverse primer is bCpGR2 (5'-AATTAACCCAAAAACTACCCTAACC-3'). The TLR9 bisulfite hemi-nested PCR is a double round PCR carried out by first amplifying the bisulfite treated DNA with the outer primers bCpGF1 and bCpGR2 in the first round PCR, and secondly using either bCpGF1 and bCpGR1 for amplification of CpG Island 1 (expected band size is 279bp) or bCpGF2 and bCpGR2 for amplification of CpG Island 2 (expected band size is 500bp) from 1ul of the initial first round PCR product. The technique is, therefore, more sensitive and specific and is capable of producing more quality amplicons especially where the starting material is low in concentration. The PCR amplification is carried out using the Robocycler (Stratagene). The 24 µl reaction mixture consists of 5 µM of forward and reverse primers and EPIK™ Amplification Kit (Bioline, London), which is a hot-start PCR kit specifically engineered to deliver highly reliable and ultra-sensitive amplification from bisulfite-modified template DNA. The reaction was carried out using the following thermocycling conditions of 10 min at 94^0C, followed by 35 cycles of 30s at 95^0C, 30s at 60^0C, 30s at 72^0C and a final cycle of 7 min at 72^0C. Following this, electrophoresis is commenced and upon confirmation of successful amplification by presence of bands after viewing under UV transilluminator, the PCR product was sent to GATC Biotech (http://www.gatc-biotech.com) for purification and sequencing.

2.11 Gel electrophoresis

Gel electrophoresis which is a standard method for separating DNA molecules of different lengths was carried out after PCR amplification of any DNA samples to allow for the necessary analytical purposes. To achieve this, 1 % agarose gel was usually prepared by weighing out 0.3g of agarose powder and placing it in a conical flask. Then 30 ml of 1 X Tris-Borate-EDTA (TBE) buffer was added to the flask which is then heated in the microwave on a maximum power for 30 s, mixed gently and heated for a further 30 s until all the agarose are melted.

The molten gel is then placed on a shaker and allowed to cool at approximately 50^0C. Then, 30 µl of diluted GEL Red (10^{-1}) is added and mixed by swirling gently before the molten gel is poured into a Perspex plate which has a comb already placed inside to form wells for PCR products and taped on the sides to prevent spillage. The gel is then allowed to set before the tape is removed and it is submerged in 1 X TBE buffer. The comb is then removed and the PCR products (10 µl PCR products mixed with 5 µl loading buffer) are loaded into the wells, together with a dye of known fragment size which monitors the electrophoresis progress. Electrophoresis is then commenced as the leads are plugged into 70V current and is completed when the dye has migrated approximately halfway down the gel. The gel is then visualised under UV transilluminator and bands of interest are purified, sequenced and analysed.

2.12 DNA sequencing

DNA sequencing of PCR products was conducted on all samples of interest after visualisation under the UV transilluminator confirmed bands of interest. After this, the PCR products and their associated primers are sent in labelled tubes to either Source Bioscience

(for trypanosome research) or GATC biotech (for TLR9 research) by post for purification and sequencing. The PCR products were sent according to their specifications. For both companies, 20 µl of PCR products (undiluted) and 20 µl of primers (10pm/ul concentration) for each sample was sent to their advised collection address. The contents of the package were sent electronically to Source Bioscience (http://www.sourcebioscience.com/) or GATC Biotech (http://www.gatc-biotech.com) and the sequence data was also received electronically.

2.13 Bioinformatics tools

A variety of computational and statistical tools are used in this research to analyse and process all biological data.

2.13.1 Finch TV

The Finch TV is a software programme that can be used to view and analyse DNA sequences. It displays raw DNA sequence data in a chromatogram viewer and is downloadable for free (http://www.geospiza.com/Products/finchtv.shtml). The Finch TV was used to analyse DNA sequences in various areas of this work (Chapter 3, 4 and 5).

2.13.2 Megablast

The Megablast is a software programme available for free on NCBI that uses input DNA sequences to search for highly similar sequences. The programme was useful in Chapter 3, 4 and 5 of this work.

2.13.3 Primer-BLAST

This tool is also available for free on NCBI and can be used to design primers or check primers for specificity. After selecting conserved regions and picking primers, the tool was

very helpful in various areas of this work (Chapter 4 and 5) for checking newly designed primers for specificity.

2.13.4 Clustal Omega

The Clustal Omega is a free tool (http://www.ebi.ac.uk/Tools/msa/clustalo/) provided by the European bioinformatics institute (EBI) for multiple DNA sequence alignments. It was used for multiple sequence alignment to identify conserved regions which were very helpful during primer design. It was also used to identify ITS Nested primer locations and this was helpful in identification of trypanosome species by deducing the expected band sizes (Chapter 3, 4 and 5).

2.13.5 EMBOSS Cpgplot

The EMBOSS Cpgplot is a free tool (https://www.ebi.ac.uk/Tools/seqstats/emboss_cpgplot/) provided by the EBI. It was used to identify CpG islands in bovine TLR9 and used to investigate TLR2, TLR4 and TLR6 for possibility of CpG islands in their sequences (Chapter 5).

2.13.6 EMBOSS needle

The EMBOSS Needle (http://www.ebi.ac.uk/Tools/psa/emboss_needle/) is a free protein alignment tool provided by EBI. It was used to determine which of the two sequences (human or mouse) is more identical or similar to bovine TLR9 reference protein (Chapter 5).

2.13.7 MethPrimer

MethPrimer is a free tool for designing bisulfite-conversion-based Methylation PCR Primers (www.urogene.org/methprimer), provided by the University of California San Francisco (Li

and Dahiya, 2002). The tool was useful in this project to select bisulfite PCR primers for regions covering the bovine TLR9 CpG Island 1 and CpG Island 2 (Chapter 5).

2.13.8 SNAP

SNAP is a tool for evaluating effects of single amino acid substitutions on protein function. It was developed at Columbia University, New York (https://rostlab.org/services/snap/). Protein sequences for comparison are entered into the programme, but the results are sent to the user via email (Bromberg and Rost 2007). SNAP was used in this project to predict the effect of non-synonymous polymorphisms found in bovine TLR9 gene on protein function (Chapter 5).

2.14 Molecular phylogenetic analysis: Single genes and concatenated sequences

The SSU-rRNA DNA sequences of 25 kinetoplastids were assembled from GenBank on NCBI database (http://www.ncbi.nlm.nih.gov/genbank/), including available LSU-rRNA sequences for 10 of the kinetoplastids. Also, the badger trypanosome SSU-rRNA and LSU-rRNA sequences derived from this study was included in the analysis. Twenty-four of the kinetoplastids were members of Trypanomastidae family, of the genus *Trypanosoma* while the remaining to be used as an out-group is a member of Bodonidae family. For single gene phylogenetic tree, the SSU-rRNA sequences were aligned with Muscle (v3.7) which achieves fast and accurate alignments (Edgar 2004) and poorly aligned sequences were removed using Gblocks, which is shown to make alignments more appropriate for phylogenetic reconstruction (Castresana 2000). This was performed on Phylogeny.fr webserver (Dereeper et al. 2008) before the selected suitable alignment was then uploaded on MEGA v. 6 (Tamura et al. 2013), where the best-fit substitution model was determined and also specified to reconstruct the phylogenetic tree based on maximum likelihood using 1000 bootstrap

replicates. Concatenated (multiple) phylogenetic tree was reconstructed using LSU-rRNA sequences of nine trypanosomes; *T. grosi* (GI: 46091659), *T. otospermophili* (GI:46091661), *T. kuseli* (GI:46091662), *T. rangeli* (662247341), *T. rotatorium* (GI: 1073254), *T. simiae* (GI:1040861), *T. congolense (riverine forest)* (GI: 1040860), *T. congolense (Kilifi)* (GI:1040858) and *T. congolense (savannah)* (GI: 1040856) including the LSU-rRNA of *Trypanoplasma borreli* (GI:1073253) and the Badger LSU-rRNA trypanosome sequence derived from this study (Accession Number: KR527480). Using the procedure described above, all LSU-rRNA sequences were aligned and ambiguous regions were removed before the alignments are added to their SSU-rRNA counterparts to derive a concatenated sequence dataset for the respective trypanosome species. Using concatenated data sets for phylogenetic reconstruction has been shown to increase the probability of phylogenetic accuracy (Gadagkar et al. 2005). Also, the alignments were uploaded in MEGA v. 6 (Tamura et al. 2013), where the best-fit substitution model was selected and the phylogenetic tree was reconstructed based on maximum likelihood using 1000 bootstrap replicates. Phylogenetic analysis was employed to determine the taxonomic status of trypanosome species detected in badgers. More details and results are available in Chapter 4 of this report**.**

2.15 Statistical analysis

Statistical analysis was carried out using IBM SPSS Statistics for Windows (version 20) (Chapter 3, 4 and 5) and also Microsoft Excel (Chapter 3 and 4).

In Chapter 3, the SPSS was used to measure the degree of agreement between single and multiple punch trypanosome detection tests. The Fishers exact test on GraphPAD software (http://graphpad.com/) was used to determine and compare prevalences of trypanosomes in

southern Nigerian cattle using two tests. The Microsoft Excel package was used forecast prevalence of trypanosome infections in southern Nigerian cattle if more tests are carried out.

In Chapter 4, Chi test in SPSS was used to measure the association between gender and trypanosome infection status of badgers. The Fisher's exact test was used to measure the association between Scaled Mass Index, setts and trypanosome infection status of badgers. The SPSS was also used to compute the odds of infection with trypanosome based on geographical region (north or south of Woodchester Park) of sett where badger live. The frequency distribution of trypanosome infection in badgers was analysed using the SPSS. Linear regression graph was plotted using Microsoft Excel package to deduce size of badger trypanosomes.

In Chapter 5, the SPSS was the main software used for statistical analysis. The frequency distribution of single nucleotide polymorphisms (SNPs) was computed with the aid of SPSS. The Chi Square test in SPSS was used to measure the association between SNPs and trypanosome infection status of cattle. The Kendall's Tau-b (nonparametric) correlation coefficient on SPSS was used to investigate the correlation between SNPs found in TLR9 mRNA of bovine samples.

Chapter 3: Prevalence of trypanosomiasis in southern Nigerian cattle

3.1 Introduction

Trypanosomes are a major cause of diseases and economic loss in Nigeria, resulting in a huge yearly loss in domestic animal production valued at about N110m (currently $552,625.70) (Nigerian Institute for Trypanosomiasis Research 2008). Clinically important trypanosome species that have been detected in Nigerian cattle include *T. b. brucei, T. vivax and T. congolense* (Enwezor et al. 2012, Omotainse et al. 2000) and the tsetse flies which infest about 80% of the Nigerian land mass are the main vectors transmitting trypanosomes in Nigeria (Abenga et al. 2004).

Cattle are reared mostly in the northern part of the country but they are usually migrated to the south due to nomadic practice involved in Nigerian cattle market (Ogunsanmi et al. 2000). Although, cattle movement was highlighted as the major cause of *T. b. rhodesiense* human sleeping sickness epidemics that occurred in some districts in Uganda (Hide et al. 1996, Picozzi et al. 2005), there has not been a reported case of *T. b. rhodesiense* HAT in Nigeria so cattle movement would not pose a direct threat for the disease epidemiology in Nigeria. However, with the possibility of host to host transmission which is now becoming very popular in trypanosomiasis epidemiology (Noyes et al. 2002, Verma et al. 2011) it becomes necessary to investigate the southern part of Nigeria for presence of trypanosomes particularly as this may also help towards contributing to sparse literature available on prevalence of trypanosomes in this region. One method that can be used generically on a wide range of trypanosome species is the Internal Transcribed Spacer (ITS) Nested PCR based on generic trypanosome primers (Ahmed et al. 2011, Cox et al. 2005, Cox et al. 2010)

but it also requires adequate sampling techniques for DNA isolation prior to PCR amplification.

Sampling using the Whatman FTA card has proven a great success for protozoa pathogenic detection and diagnosis using PCR (Ahmed et al. 2011, Ahmed et al. 2013, Cox et al. 2005, Cox et al. 2010). However, there are concerns over determining the true prevalence of trypanosome infection from a sample using single punches of FTA cards for direct PCR amplification, as lower than expected prevalence have been estimated in some studies after its application (Adams et al. 2008, Cox et al. 2010). The major factor contributing to the decreased sensitivity and prevalence is the uneven distribution of *Trypanosoma* DNA in the FTA card matrix (Cox et al. 2010). One efficient approach of investigating trypanosome infection from a sample of FTA cards is through using Chelex® resin to elute *Trypanosoma* DNA from an already prepared FTA cards. Ahmed et al. (2011), demonstrated that when using Chelex® resin (for the elution of trypanosome material), the sensitivity using 10 eluted whole blood punches increased significantly from 35.6% to 56.4%, compared to non-eluted whole blood using 10 separate discs.

Additionally, although the simplicity of using FTA cards for nucleic acid isolation, storage and processing is an added prominence for the molecular epidemiologist, the process of preparing the FTA cards would require repetition when trypanosome has been detected from a single punch and the researcher decides to go back to the same FTA card to re-amplify another single punch in the hope of detecting the same *Trypanosoma* species from the sample. This can be time-consuming and often involves the re-use of laboratory resources which could be financially demanding. The possibility of detecting trypanosome from the next FTA card preparations sometimes may not be guaranteed, although what has been shown to increase the likelihood of detecting trypanosomes from each sample of FTA cards is by preparing several punches from the same FTA card sample (Cox et al. 2010). These

processes would seem laborious so it becomes crucial for the researcher to preserve any strain of trypanosomes that have been successfully detected by PCR amplification from FTA cards, especially if it requires re-amplification.

One particular method that has proved a success in cytology is Touch Preparation (Touch prep). It involves obtaining a single core biopsy sample from the core material for diagnosis. When performed carefully and correctly Touch Prep is reported to offer rapid diagnosis whilst simultaneously preserving the core material for subsequent use (Hahn et al. 1995). Its application is becoming very popular in medical sciences. In another study, (Khamechian et al. 2012) showed how by using touch prep technique smears taken from patients with brain tumors using grease-free glass slides and subsequently stained with Giemsa and Papanicolaou can accurately and easily be diagnosed by microscopic examination. Their work suggests it is less time and also effective compared to the rigorous procedure of waiting till after surgeries before processing the remaining tissues for the diagnosis. Thus, it was hypothesised for our study that in a similar vein we could develop and apply a Gel-based touch preparation technique to preserve the DNA of trypanosomes detected from FTA cards by ITS-Nested PCR to use for subsequent PCR amplification.

Hence, the purpose of this study was to apply the ITS Nested PCR (Cox et al. 2005) for detection of trypanosomes in southern Nigerian (Ahoada) cattle whilst simultaneously developing an Agarose Gel-based Touch Preparation technique to enable a less laborious PCR re-amplification of any trypanosome strain initially detected. The following objectives were set out;

- Application of the ITS-Nested PCR to diagnose trypanosomes from single punches taken from FTA cards.

- Application of the ITS-Nested PCR to diagnose trypanosomes from multiple punches taken from FTA cards.
- Analysis and comparison of prevalence between the single and multiple punch extraction methods.
- Development and application of the Agarose Gel-based Touch Preparation technique to preserve stocks of trypanosome positive PCR products.

3.2 Materials and methods

A total of 80 bovine blood samples obtained from bulls in Nigeria were available for testing (Table 3.1).

Table 3.1. Sample codes for 80 bulls.

Sample Codes									
AH1	AH2	AH3	AH4	AH5	AH6	AH7	AH8	AH9	AH10
AH11	AH12	AH13	AH14	AH15	AH16	AH17	AH18	AH19	AH20
AH21	AH22	AH23	AH24	AH25	AH26	AH27	AH28	AH29	AH30
AH31	AH32	AH33	AH34	AH35	AH36	AH37	AH38	AH39	AH40
AH41	AH42	AH43	AH44	AH45	AH46	AH47	AH48	AH49	AH50
AH51	AH52	AH53	AH54	AH55	AH56	AH57	AH58	AH59	AH60
AH61	AH62	AH63	AH64	AH65	AH66	AH67	AH68	AH69	AH70
AH71	AH72	AH73	AH74	AH75	AH76	AH77	AH78	AH79	AH80

All the samples were of *Bos indicus* subspecies, and a major feature for identifying these species is the possession of a humpback (Loftus et al. 1994). The samples used in this study were of northern origin but were transported down south by herdsmen. The samples were

collected by butchers from an abattoir in Ahoada, a south-south town in Rivers State (Fig. 3.1). In the abattoir very few cows are slaughtered per week, which could be due to the high cost of purchasing a cow. So, given the short time available for sample collection (2 weeks), only 80 samples were successfully collected.

Figure 3.1. Map of Nigeria. Black arrow down south shows location of abattoir in Ahoada, Rivers state.

The photos captured by the butchers confirm the bulls to possess humpbacks indicating they are *B. indicus* species (Fig. 3.2).

Figure 3.2. Photos of White Fulani bulls used in this study. The humpback feature present indicates they are *B. indicus* species. Photo captured by an Ahoada butcher named Happy.

The blood from each sample was collected and stored on Whatman FTA and prepared using the protocol described in Chapter 2.2.2 of this report. After preparation of FTA cards, a direct ITS-Nested PCR (Chapter 2) was applied to a prepared punch of FTA card from each sample for diagnosis trypanosomes. Multiple punches were collected from each individual FTA cards

and the bovine DNA was eluted using Chelex ® elution protocol (Chapter 2.4). PCR products were analysed after gel electrophoresis to determine band sizes obtained and this was used to identify the species of trypanosome detected in samples using the estimated species-specific sizes pointed out by (Cox et al. 2005). Lastly, the prevalences between the ITS-Nested PCR on single punch prepared sample and ITS-Nested PCR on Chelex eluted DNA from multiple punches were compared using Chi-square (Goodness of fit) test. The Kappa coefficient analysis was then applied to measure the degree of agreement between the two tests.

3.3 Results

3.3.1 ITS-Nested PCR: Single punch

This objective was targeted at determining if the bovine DNA samples from FTA cards would amplify with the ITS primers and thus identify the primers were detecting trypanosomes in the samples. A single punch of 1mm disc was randomly taken from an FTA® card and used in the first round of the nested PCR and 1µl of that product was used in the second round PCR. The procedure for the ITS-Nested PCR is described in Chapter 2.4. Ten out of the 80 samples prepared from a single punch and amplified using the ITS-Nested produced band sizes indicating they were positive for trypanosomes (Fig. 3.3) with a resulting 12.5% prevalence rate (6.7%-22.6%, 95% CI).

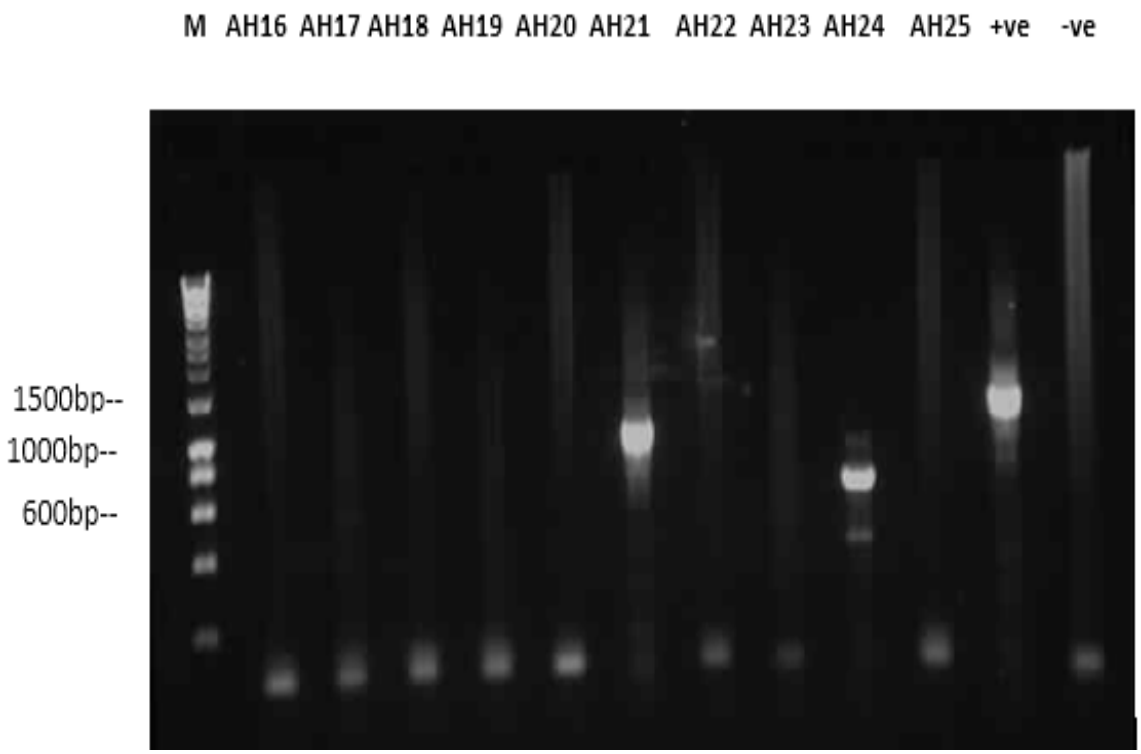


Figure 3.3. Gel image showing detection of trypanosomes in Ahoada samples. Lane M represents the marker. Lane AH16- AH25 are samples investigated for trypanosomes using ITS-Nested PCR. Bands on lane AH21 and AH24 indicated the samples were positive for trypanosomiasis. AH24 is infected with mixed (3) species of trypanosomes. +ve (*T. b. brucei*) and –ve are the control samples.

It can be recalled that the ITS Nested PCR produces a unique band size for each species of trypanosome and is capable of identifying mixed species infection. The size of most clinically important African trypanosomes affecting Nigerian cattle were deduced and specified in the work of (Cox et al. 2005). In this study, the NCBI database was accessed and the exact ITS primer positions for species of trypanosomes and estimated band sizes using any set of the ITS-nested primers are presented below (Table 3.2). This served as a guide towards identifying the species of trypanosomes detected in these samples.

Table 3.2. Some trypanosomes and their expected band sizes using ITS-Nested primers

Accession number (Trypanosome species)	ITS1 Primer Location	ITS2 Primer Location	ITS3 Primer Location	ITS4 Primer Location	Expected Band size with ITS1 and ITS 2	Expected Band size with ITS3 and ITS4
AF 306776.1 ***T. brucei***	X	X	73	1289	X	1216
AF 306777.1 ***T. brucei***	X	X	73	1285	X	1212
U 22318 ***T. (tsavo) Congolense***	2057	3439	2178	3131	1383	954
U 22320 ***T. Simiae***	2059	3337	2180	3029	1279	850
AB007814.1 ***T. theileri***	X	X	2153	3142	X	990
JN673395.1 ***T. theileri***	X	X	1	930	X	930
U 22316 ***T. vivax***	1970	3003	2090	2700	1034	611
U 22317 ***T. Congolense***	3558	5412	3679	5100	1855	1422
U 22315 ***T. Congolense***	3136	4981	3257	4669	1846	1523
U 22319 ***T. Congolense***	3025	4970	3146	4658	1946	1513
AB175625.1 ***T. Otospermophili***	2067	3719	2188	3392	1653	1205

** Accession numbers in the table are provided by NCBI for access to DNA sequences of the trypanosomes. Expected band size is the difference between the ITS-Nested primer positions. Those in red are sizes after amplification with ITS3 and ITS4. X in table implies that region is unavailable on NCBI database, making it impossible to deduce expected band size.

Based on this, band sizes from PCR products were estimated following visualisation under UV light after gel electrophoresis and the species of trypanosomes detected on Single punch amplified using ITS-Nested PCR are shown in Table 3.3 below. As shown in the table, the trypanosome species identified were *T. theileri, T. vivax* and *T. simiae*, with estimated band sizes of 930/990, 611, and 850 base pairs (bps), respectively. Fig. 3.3 shows samples AH24 to be infected with mixed species, these appeared as triple bands on the gel image and the species thus identified using ITS-Nested PCR on the single punches for all samples are listed below (Table 3.3).

Table 3.3. Trypanosomes detected using ITS-Nested on single punches

	Single punch on FTA®	
Samples	**Estimated Band Size (bp)**	**Estimated Trypanosome spp**
AH 11	990	*T. theileri*
AH 21	990	*T. theileri*
AH 24	611/988/550	*T. vivax/T. theileri/* Unknown
AH 26	611	*T. vivax*
AH 27	611	*T. vivax*
AH 28	611	*T. simiae*
AH 29	990	*T. theileri*
AH 30	850	*T. vivax*
AH 38	611	*T. vivax*
AH 41	611	*T. vivax*

Furthermore, the PCR products of four of the samples that tested positive for trypanosomes (AH11, AH21, AH38 and AH41) were sent to a commercial sequencing company (Source Bioscience) for purification and sequencing. The sequence data received from the company was opened using a DNA sequencing analysing software (Finch TV) and the Megablast software programme from the National Center for Biotechnology Information (NCBI) was used to identify homologous DNA sequences in the NCBI databases.

The sequence data showed partial alignment to the ITS region of trypanosomes in all samples. AH11 and AH21 initially identified at *T. theileri* based on their estimated band size (990bp) were shown to be 99% identical to *T. theileri* on the database which confirms their identity (Fig. 3.4) but AH38 and AH41 identified as *T. vivax* based on their band size did not produce good alignments, they only managed to align partially with some species of trypanosomes so their identity could not be established using this method. The difficulty in obtaining good sequences using the ITS-Nested primers may be partly due to positions of the primers on the ITS gene, particularly as following these positions are the variable ITS region

of trypanosomes. More sophisticated sequencing methods such as next generation sequencing might be useful in obtaining sequences covering this region.

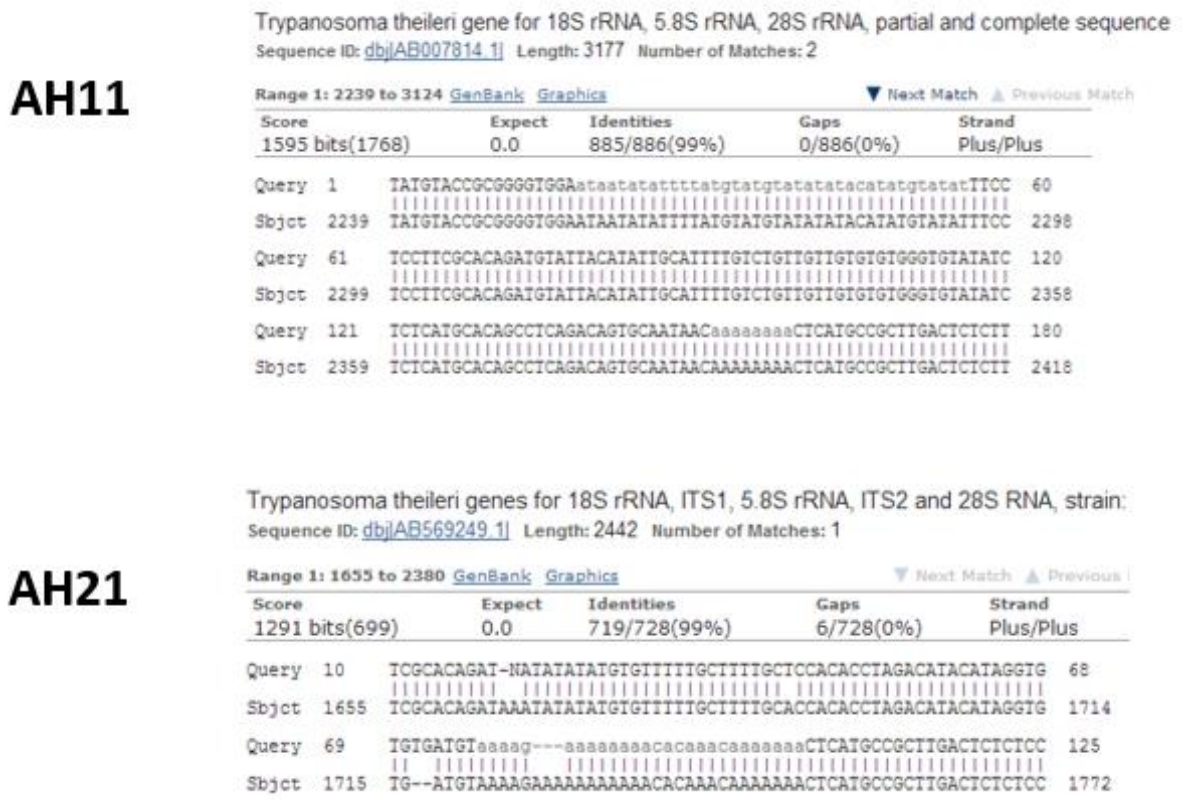


Figure 3.4. Alignments of AH11 and AH21 with *T. theileri*. AH11 above shows 99% similarity with *T. theileri* and this is same as AH21.

3.3.2 ITS-Nested PCR: Multiple punches (Chelex® resin extraction)

The objective of this study was to determine if the application of ITS-Nested PCR to detect trypanosomes from DNA samples eluted using Chelex® resin extraction from multiple punches of FTA cards would increase the prevalence of trypanosome infection as compared to the samples amplified using the ITS-Nested PCR from single punches. Multiple punches (7 punches) of 1mm discs were randomly taken from an FTA® card. The FTA cards were prepared using FTA purification agent as described in Chapter 2.2.2 and the DNA was eluted using a Chelex® resin extraction procedure (Chapter 2.2.3). Following this, 1µl of the eluted

DNA was used in the first round reaction of the ITS-Nested PCR then another 1µl of the first round PCR reaction product was added to a second round mixture. These reaction procedures for the ITS-Nested PCR are elaborated in Chapter 2.4. The eighty samples of the *Bos indicus* cattle were available for the study. Seventy-one (71) of these, where tested using this procedure and 9 resulted in successful amplification of the ITS region of trypanosome producing varying band sizes indicating the samples were infected with different species/subspecies of trypanosomes (Fig. 3.5). This gives a prevalence of 12.7% (6.6%-22.6%, 95% CI). MSc Student Tanya Lepore assisted with this work.

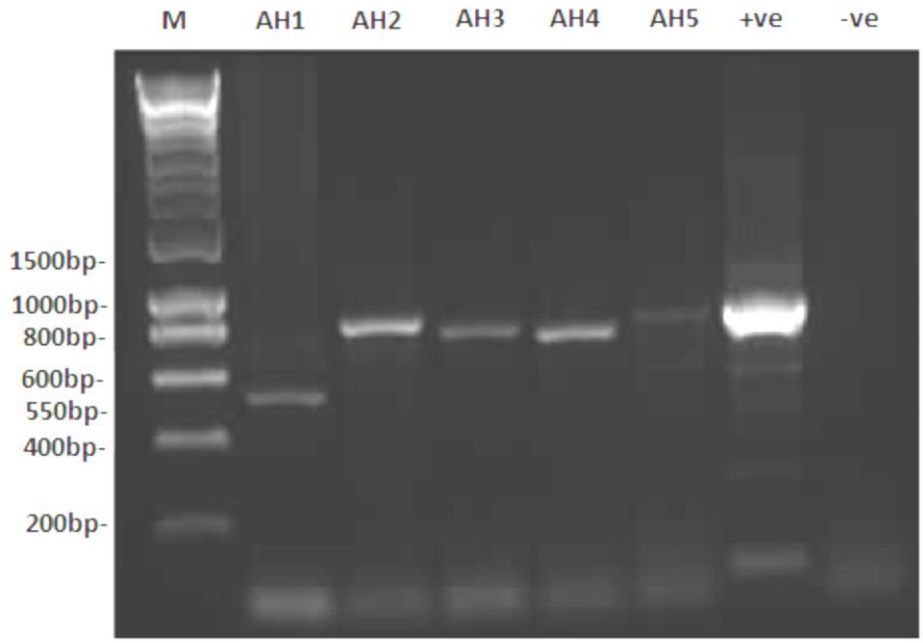


Figure 3.5. ITS-Nested PCR amplification after Chilex resin extraction. (+ve) represents positive control, using *Trypanosoma avium* (bird trypanosome*)* and (-ve) represents the negative control, using distilled water. Sample AH1 was shown to be positive for an unknown species, *T. vivax, and T. simiae,* AH2, AH3, AH4 were positive for *T. simiae* and AH5 was positive for an unknown species, *T. vivax* and *T. theileri.* M represents HyperLadder 1kb marker.

As shown in Fig. 3.5, AH1 presented with varying band sizes of around 550 base pairs (bps), 611 bps and 850 bps, AH2 had a band size of 930bps, AH3 and AH4 showed band sizes of around 850 base pairs while AH5 had band sizes of around 611 and 930 base pairs. An

unidentified species was shown to have an estimated band size of 550 bps (Fig. 3.5). All trypanosomes were identified based on their band sizes (Table 3.4).

Table 3.4. Trypanosome species detected using ITS-Nested PCR on Chelex® resin eluted DNA from multiple punches.

Chelex® resin		
Samples	**Estimated Band size (bp)**	**Estimated *Trypanosoma* spp**
AH1	550/620/850	Unknown/*T. vivax*/*T. simiae*
AH2	930	*T. theileri*
AH3	850	*T. simiae*
AH4	850	*T. simiae*
AH5	611/930	*T. vivax*/ *T. theileri*
AH10	850	*T. simiae*
AH11	611	*T. vivax*
AH25	550	Unknown
AH30	550/850/990	Unknown/*T. simiae*/ *T. theileri*

3.3.3 Analysis of trypanosome prevalence

Analysis of both methods (extraction from single punch and extraction from multiple punches using chelex resin) shows that a total of 17 out of 80 samples amplified using ITS-Nested PCR tested positive for trypanosomes, resulting in an overall prevalence rate of 21.3% (13.6% - 31.5%; 95% CI) (Table 3.5). Three species of trypanosomes were identified in the samples based on their band sizes after electrophoresis (Cox et al. 2005) namely; *T. simiae* (6.25%), *T. theileri* (8.75%), and *T. vivax* (12.5%). Additionally, unknown species of trypanosomes (5%) were also detected in some samples while some were infected with mixed species of trypanosomes (5%) (Fig. 3.6; Table 3.5).

Table 3.5. Trypanosome positive samples

Sample code	Infected (Y/N)	Estimated species & band size (bp)	Sample code	Infected (Y/N)	Estimated species & band size
AH1	y	Unknown (550bp)	AH41	y	*T. vivax* (611)
AH2	y	*T. theileri* (930)	AH42	n	
AH3	y	*T. simiae* (850)	AH43	n	
AH4	y	*T. simiae* (850)	AH44	n	
AH5	y	*T. theileri/ T. vivax* (930, 611)	AH45	n	
AH6	n		AH46	n	
AH7	n		AH47	n	
AH8	n		AH48	n	
AH9	n		AH49	n	
AH10	y	*T. simiae* (850)	AH50	n	
AH11	y	*T.vivax/ T. theileri* (611, 990)	AH51	n	
AH12	n		AH52	n	
AH13	n		AH53	n	
AH14	n		AH54	n	
AH15	n		AH55	n	
AH16	n		AH56	n	
AH17	n		AH57	n	
AH18	n		AH58	n	
AH19	n		AH59	n	
AH20	n		AH60	n	
AH21	y	*T. theileri* (990)	AH61	n	
AH22	n		AH62	n	
AH23	n		AH63	n	
AH24	y	Unknown/*T. vivax/T. theileri* (550, 611, 930)	AH64	n	
AH25	y	Unknown (550)	AH65	n	
AH26	y	*T. vivax* (611)	AH66	n	
AH27	y	*T. vivax* (611)	AH67	n	
AH28	y	*T. vivax* (611)	AH68	n	
AH29	y	*T. theileri* (990)	AH69	n	
AH30	y	Unknown/*T.vivax/ T. simiae/T. theileri* (550, 611, 850, 990)	AH70	n	
AH31	n		AH71	n	
AH32	n		AH72	n	
AH33	n		AH73	n	
AH34	n		AH74	n	
AH35	n		AH75	n	
AH36	n		AH76	n	
AH37	n		AH77	n	
AH38	y	*T. vivax* (611)	AH78	n	
AH39	n		AH79	n	
AH40	n		AH80	n	

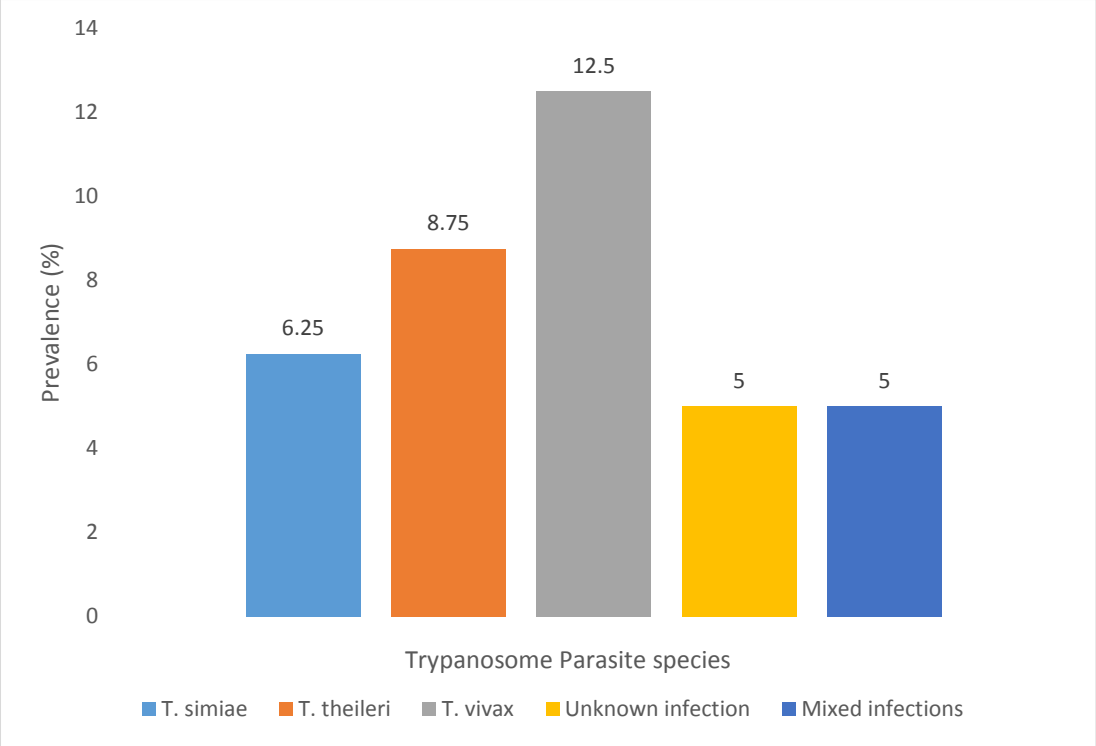


Figure 3.6. Bar chart showing overall prevalences of trypanosome species detected in Nigerian White Fulani cattle in Ahoada using both, single punch and multiple punches DNA preparation techniques prior to amplification using ITS-Nested PCR. Colour legends in the chart represent the names of trypanosome species while numbers represents percentages prevalences.

As shown in Fig. 3.6 above, the majority of samples were infected with *T. vivax* (n=10, 12.5%) and *T. theileri* (n=7, 8.75%). A prevalence of 6.25% (n=5) was recorded for infection with *T. simiae* while 5% prevalence (n=4) was seen in both unknown and mixed infections. Also, 3 out of the 4 mixed infections and 3 out of the 4 unknown infections were detected from DNA amplified by ITS-Nested PCR after using the Chelex resin extraction technique.

Furthermore, the prevalence of positively diagnosed *Trypanosoma* species for using the single punch on FTA® cards was calculated as 12.5% while the prevalence using Chelex® resin was calculated as 12.7%. Although, by mere comparison the prevalence increased by 0.2% but by using the Fishers exact test on GraphPad software (http://graphpad.com/) to compare the difference in both prevalences, it was shown that there was no significant

difference between the first test (diagnosing trypanosomes from Single Punch taking from FTA cards) and the second test (diagnosing trypanosomes from seven punches taken from FTA cards) as the P value equals 1.000. The Kappa Coefficient statistics on SPSS was used to measure the degree of agreement between the two tests. According to Landis and Koch (1977), the standards for strength of agreement for Kappa coefficient are; ≤0 = poor, 0.1-0.20 = slight, 0.21-0.40 = fair, 0.41-0.60 = moderate, 0.61-0.80 = substantial, 0.81-1 = almost perfect. The Kappa's coefficient value from our measurement using the SPSS is 0.089, indicating a poor level of agreement between the two tests.

3.3.4 Agarose gel-based touch preparation technique

The objective of this study was aimed at developing and applying a gel-based touch preparation (touch prep) technique to preserve strains of trypanosomes detected from FTA cards by ITS Nested PCR to use for subsequent PCR amplification. The ITS-Nested PCR was carried out on ten Ahoada samples (AH31-AH40), with *T. brucei* used as a positive control. After electrophoresis, the gel was viewed under UV light. Two bands were shown in AH38 and *T. brucei* control sample lane on the gel indicating the samples were positive for trypanosomes (Fig. 3.7). The sample AH38, as shown on the gel had an estimated band size of 620bp indicating it was infected with *T. vivax* in reference to Cox et al. (2005) specifications for species identification using ITS-Nested PCR.

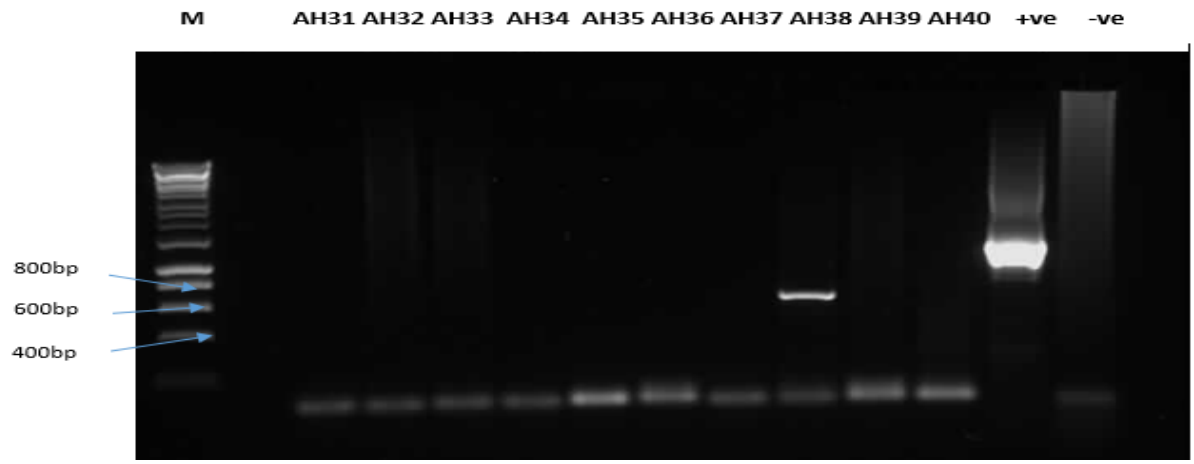


Figure 3.7. Gel image showing trypanosome detected from FTA card using ITS Nested PCR. AH 38, is a sample positive for *T. vivax* with an estimated band size of 620bp. All other AH samples were trypanosome negative since there were no visible bands on the gel. Lane M represents the marker. +ve represents *T. brucei* sample used as control and –ve represents water as a control samples.

The gel was then transferred to a dark UV room where the gel-based touch prep was carried out. Materials required to commence the technique was the gel containing the positive sample, UV light, a sterile 10ul tip and 20ul of PCR graded H_2O contained in an Eppendorf tube. The gel was placed on the slide where the UV light can reflect to show clearly the bands of interest. With the lab coat and goggles worn prior to proceeding with the technique, the UV transilluminator was then switched on. When the band of interest is unveiled, the sterile tip was then used to make several punches on the visible bands from sample AH38 (620bp, positive for *T. vivax*) and Control sample (1216bp, positive for *T. brucei*) while at the same time ensuring that the tip was not touching the surface of the slide while performing the punch. The tip was then placed in the Eppendorf tube containing the PCR graded water and allowed to settle for 10 minutes. Following that, the tip was then discarded and the Eppendorf tube contained the DNA which was stored at 4 ℃ or -20 ℃ for long-term storage.

ITS PCR was then carried out using the second set of ITS-Nested PCR primers (ITS3 and ITS4) in just a single round PCR according to the reaction cycling conditions detailed in

Chapter 2. 1ul of the template DNA from touch prep of control sample (*T. brucei*, 1216bp) was added to a PCR reaction mixture, 1ul each of template DNA from touch prep of AH38 (*T. vivax*, 620) was added to three other tubes containing ITS PCR reaction mixture while 1ul of water was added to the final tube containing PCR reaction mixture. After PCR amplification using ITS3 and ITS4, gel electrophoresis was performed and the gel was viewed under UV light according to the procedures described in Chapter 2 of this report.

The result showed very clear bands with the expected estimated sizes (*T. brucei,* 1216bp and *T.vivax*, 620bp) obtained in the initial ITS-Nested PCR (Fig. 3.8), indicating the agarose gel-based touch prep works as an effective tool to reamplify DNA of trypanosomes after amplification.

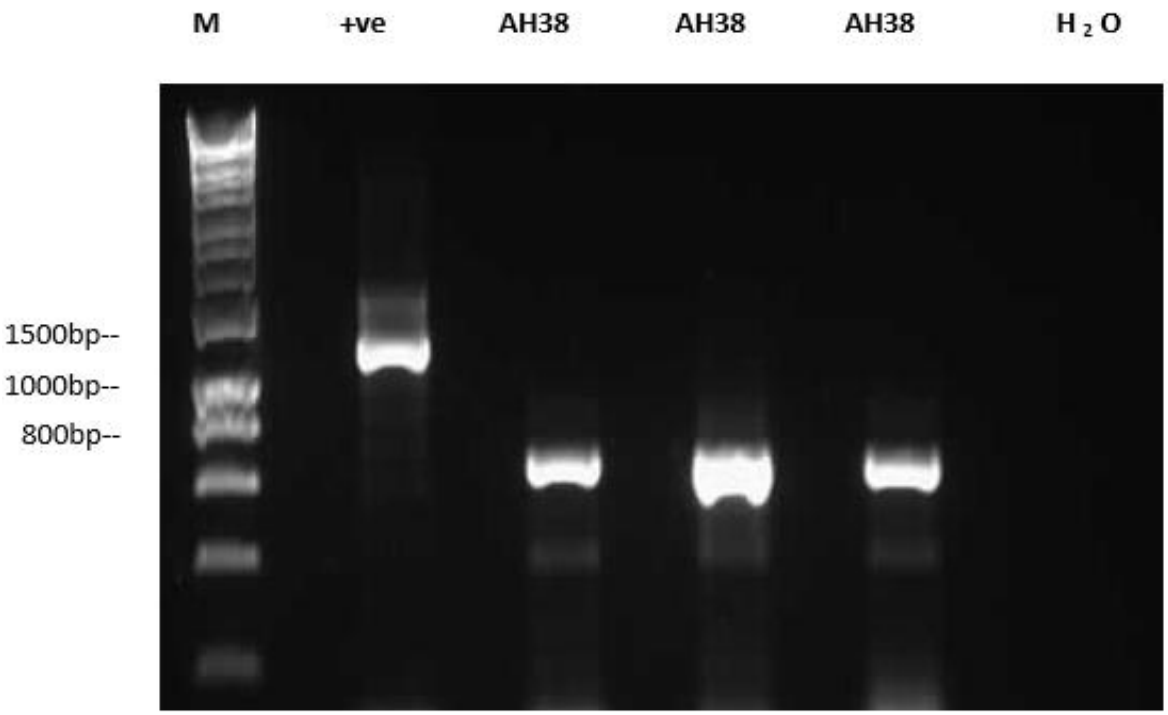


Figure 3.8. Agarose gel-based Touch Prep image from AH38 and *T. brucei* sample using ITS PCR. Lane M represents the marker. +ve is the control sample positive for *T. brucei* (1216bp) while AH38 sample lanes are positive for *T. vivax* (620bp).–ve is water used as a control.

The technique was also tested on trypanosomes detected in a different mammalian host. *T. pestanai* from British badger (sample 12K) was used as a control in one of the ITS-Nested PCR applied to investigate trypanosomes in the Ahoada cattle sample population (AH26-AH30). The results indicated sample AH29 was positive for *T. theileri* with an estimated band size of 990bp and the control samples (12K and 2848) were also positive for *T. pestanai* (1271bp) and *T. brucei* (1216bp) respectively (Fig. 3.9).

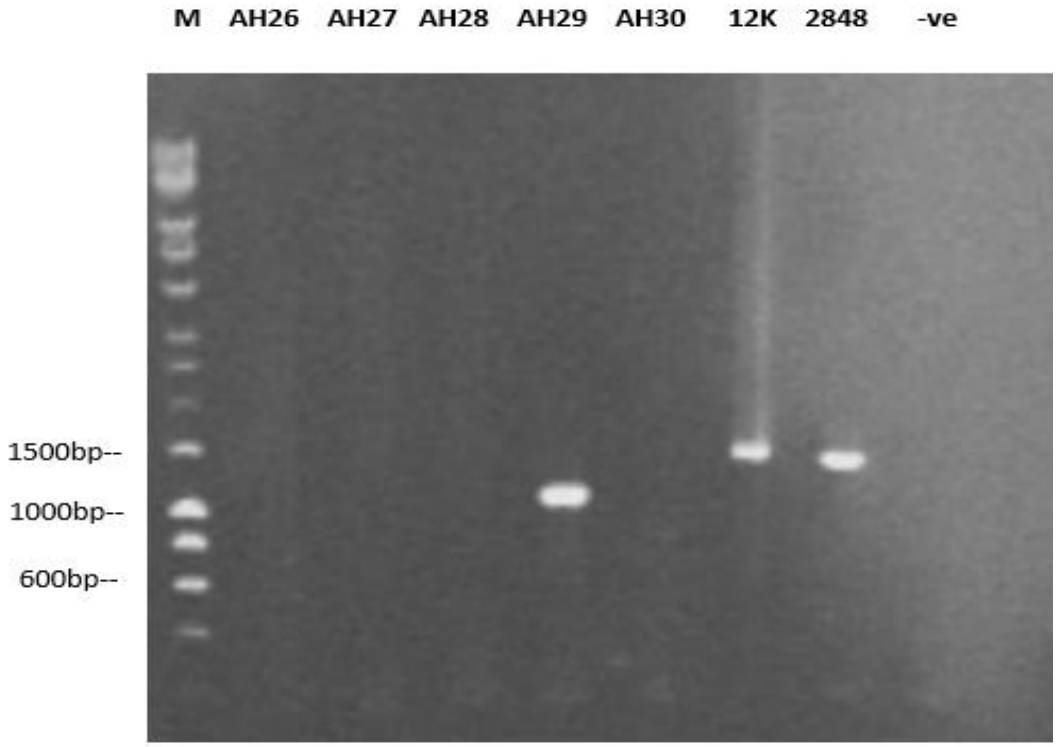


Figure 3.9. ITS-Nested PCR amplification of AH26-AH30 using *T. pestanai* and *T. brucei* as control samples. Sample AH29 is positive for *T. theileri* (990bp). Sample AH26, AH27, AH28 and AH30 were all negative. Both 12K and 2848 are control samples positive for *T. pestanai* (1271bp) and *T. brucei* (1216bp) respectively.

The newly developed agarose gel-based touch preparation technique was then applied to Badger 12K control sample of 1271bp positive for *T. pestanai* by using methods explained above. The results following gel electrophoresis showed clear bands of the same size with *T.*

pestanai (1271bp) which indicated the technique also works on trypanosomes detected from a different mammalian host (Fig. 3.10).

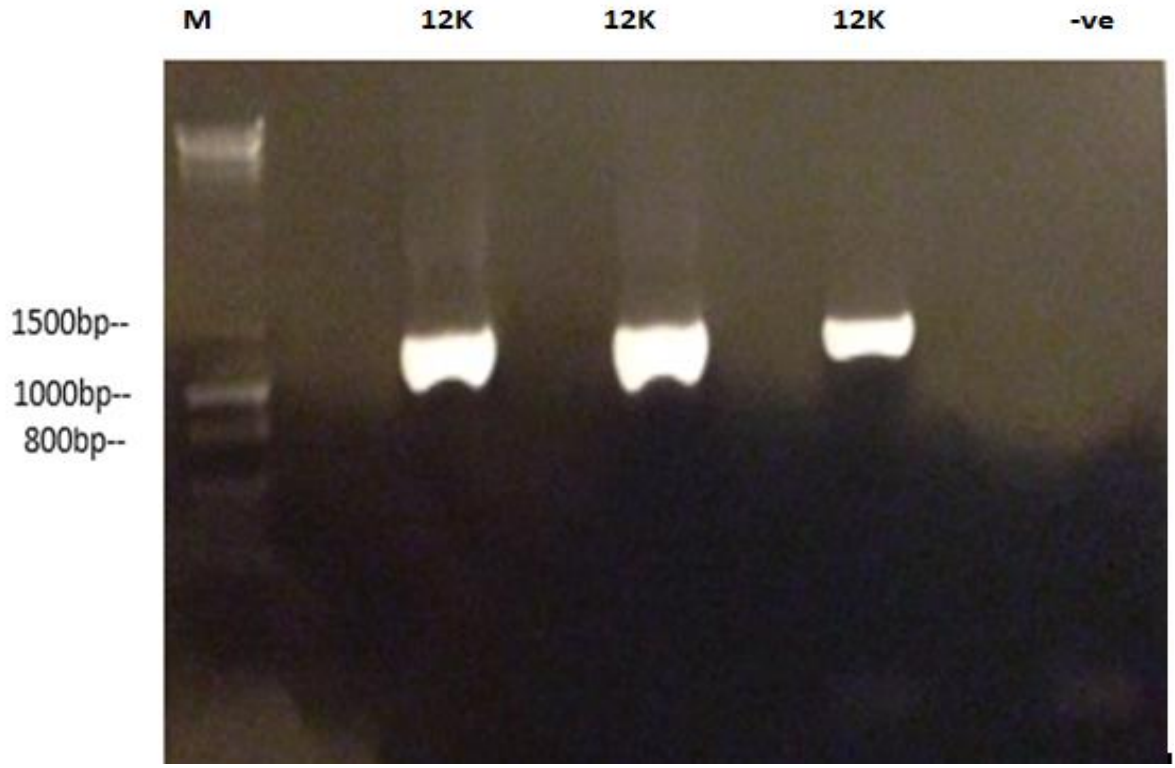


Figure 3.10 Agarose gel-based touch prep of badger 12K sample. Estimated bands sizes of 1271bp visible on 12K lanes is an indication the technique had worked on a trypanosome detected in a different mammalian host. Lane M is hyperladder 1kb plus marker of known molecular weights.

The result also indicates that on any trypanosome detected from a sample of FTA card using the ITS Nested PCR, we can avoid a subsequent laborious stage of FTA card preparation and ITS Nested PCR to detect the same strain of trypanosome in the FTA card sample through employing the gel-based touch preparation technique and running a one round PCR with the associated primers. This would enable us save time and could be a more cost effective way of preserving the trypanosome strains compared to than the usual purification and preservation kits.

3.4 Discussion

This study records an overall 21.3% (13.6% - 31.5%; 95% CI) prevalence of trypanosomiasis infection from *B. indicus* bulls sampled in Ahoada, a southern town of Nigeria (Table 4.5). Previous studies (Abenga et al. 2004, Enwezor et al. 2009) have shown lower prevalence rates (9.1% and 8.4% respectively) compared to our study but this may be due to their approach to diagnosing bovine trypanosomiasis using traditional parasitological and immunological methods. Molecular based PCR techniques offer a much more sensitive approach for diagnosing trypanosomes. For example, Takeet et al. (2013) used microscopy and PCR methods to survey pathogenic trypanosome infection in naturally infected Nigerian cattle. From the 411 cattle sampled, microscopy resulted in 15.1% prevalence while PCR resulted in 63.7% prevalence. Although, their prevalence of 63.7% using PCR was higher than the prevalence of 21.3% reported in this study, this difference may be due to their method of DNA extraction. Their study examined DNA extracted from whole blood using kits while our study examined DNA prepared from blood spotted FTA cards. Determining maximum prevalence from FTA cards is dependent upon the number of punches used for the PCR reaction. Lower than expected prevalences have been reported in some studies that used FTA cards for sampling (Adams et al. 2008, Cox et al. 2010) and uneven distribution of the parasite on the FTA card matrix was pointed as the major factor causing the decreased sensitivity and low prevalence (Cox et al. 2010). Hence, the more the punches are taken from a sample of FTA card, the higher the chances of detecting the parasite. A comparison of this likelihood was demonstrated in this study. The single punch ITS-Nested PCR resulted in 12.5% prevalence which was 0.2% lower than the prevalence of 12.7% obtained using the Chelex resin (multiple punches) ITS-Nested PCR. However, using the fisher's exact test this difference was shown to be not statistically significant (P=1.000, FET). Also, the Kappa coefficient value for measurement of agreement between the two tests is 0.089 which

indicates a poor agreement (Landis and Koch 1977). The findings from comparing these two methods in our study are not in consonance with the report by Cox et al. (2010) who demonstrated how overall prevalence of trypanosome infection in their study population increased from 9.7% (using the single punch) to 86% (using repeated/multiple punches). However, a direct comparison can't be made with their study as more punches were taken out from the FTA cards and each sample was subjected to between 92 and 114 individual PCR assays. Perhaps if more punches were taken out of the FTA cards used in this study and examined using the sensitive ITS-Nested PCR then the prevalence is likely to increase.

Three species of trypanosomes (*T. vivax, T. theileri and T. simiae*), an unknown species and mixed infections were detected using the ITS-Nested PCR in this study. The species were mostly identified based on their band sizes (Table 3.2). *T. vivax* was recorded as having the highest prevalence (12.5%) in this study. This corresponds with reports from other studies conducted in Nigeria that also identified *T. vivax* as having the highest prevalence (Anene et al. 1991, Daniel et al. 1993, Kalu 1995, Omotainse et al. 2000). Takeet et al. (2013) also reported high prevalence (23%) of *T. vivax* in southern Nigerian cattle and elsewhere in Uganda a high prevalence of 22.9% for *T. vivax* has also been reported (Cox et al. 2010) Also, this study records 8.75% prevalence of *T. theileri* (Fig. 3.6). One of the earliest documented reports of the parasite's detection is backdated down to early 1930's, a time when the parasite was even thought to be cryptic and revealed only through culture methods (Haore 1972). Unfortunately, literature on the prevalence of this parasite in Nigeria cattle is still sparse. This may be due to the non-pathogenic nature of the parasite which perhaps presents it a less clinically important species, thereby resulting in lack of interest in its investigation. The sparse literature may as well be due to the kind of methods usually applied to diagnose trypanosomiasis in Nigeria. For example, some studies base their diagnostic approaches on parasitological and immunological methods (Abenga et al. 2004, Enwezor et

al. 2009). While this is useful for the diagnosis of trypanosomiasis, it is a less sensitive approach compared to molecular-based approaches (Takeet et al. 2013). For studies that applied molecular-based approaches their primers are often designed to be species-specific targeting mainly clinically important trypanosomes for PCR amplification (Majekodunmi et al. 2013, Takeet et al. 2013). Therefore, there methods do not offer the possibility of detecting other species of trypanosomes such as *T. theileri* thus introducing potential bias to their studies. Using a method such as ITS-Nested PCR offers a more unbiased approach for diagnosing trypanosomiasis. For example, *T. theileri* was detected in this present study population. By using the ITS-Nested PCR, *T. theileri* was detected and found to have the highest prevalence rate (85.7%) in research carried out by Cox et al. (2010). This study also records a 6.25% prevalence of *T. simiae* infection (Fig. 3.6). Although the cattle is an unusual host for the parasite, early records of detection of the parasite from cattle in Nigeria goes as far back early 1950's and 1960's as highlighted in the work of Haore (1972). Literature on infection of livestock in Nigeria with the parasite is still sparse but in other parts of Africa *T. simiae* have been detected in Tsetse flies from Kenya (Njiru et al. 2004) and domestic ruminants in Cameroon (Nimpaye et al. 2011). Also, prevalence of *T. simiae* in Nigeria may be underestimated based on the less sensitive (microscopy) and species-biased (species-specific) methods usually applied to investigate the prevalence trypanosomiasis in Nigeria. Also, since the general morphology of *T. simiae* is similar to that of *T. congolense* (Haore 1972) issues of inaccurate species identification when using lesser sensitive methods cannot be ruled out. A major strength of the ITS-Nested PCR is its capability of detecting mixed infections and unknown trypanosome species even when the parasitaemia is very low. From this study, a prevalence of 5% was recorded for both samples infected with mixed infections and samples infected with unknown species. Interestingly, ITS-Nested PCR using single punch of FTA card accounted for 1.25% (n=1) of the prevalence observed in both mixed

infections and unknown infections while ITS-Nested PCR using chelex resin (multiple punches) accounted for the remaining 3.75% (n=3). These prevalences are likely to increase following repeated tests using more punches.

One limitation of the ITS-Nested PCR is its inability to always identify species of trypanosomes despite sensitively detecting them. In most cases this would be due to the absence of the ITS regions for some/new species of trypanosomes on the databases, thereby making it impossible to search the databases for homologous sequences using DNA sequences derived from the new species. However, with the steady influx of genes into the databases nowadays such challenges could soon be subdued.

Additionally, the FTA card is emerging a useful sampling tool based on simplicity in processing, storage and transportation of biological samples. The development of the agarose gel-based touch preparation from this study could prove a cost-effective way of preserving trypanosome DNA from bands shown on gels compared to the usual purification and preservation using kits. It could also be useful when DNA quantity is low since re-amplification with the process increases the amount of DNA. In terms of trypanosomes, it could also be useful to isolate various strains of the parasites in the case of mixed infections or in methods involving multiplex PCR.

For example, Picozzi et al. (2008) developed a multiplex PCR based on amplification of trypanosome SRA and *GPI-PLC* genes. Amplification using their technique, therefore, produces two bands- one for the SRA gene and the other for the *GPI-PLC* gene. Where, for instance, an investigator applies their technique and then it becomes necessary to isolate DNA from any of the two PCR amplicons for perhaps sequencing purposes or re-amplification, the agarose gel-based technique might be useful to overcome such challenges. One limitation of the agarose gel-based technique is that DNA preserved after the process can

only be re-amplified using the same primers that produced the bands shown on gel initially before the process or using different primers that are with the internal region of the initial primers used for amplification. This limits the researcher on what can be done with the available DNA.

In conclusion, the results of this study show that the ITS-Nested works as an effective tool for diagnosing various species of trypanosomes. It also shows that trypanosome infections continue to be a major cause of disease in Nigerian cattle. It identifies *T. vivax* as having the highest prevalence in study population. It suggests the use of repeated testing and punching from various areas of the FTA card matrix when used as a sampling tool for diagnosis of trypanosomes. Lastly, it presents a newly developed agarose gel-based touch preparation technique as a cost-effective way of preserving trypanosome strains from gels and suggests the method could be applied generally to preserve DNA stock from gel bands of interest.

Chapter 4: Detection of trypanosomes in the Eurasian badgers using ITS-nested PCR

4.1 Introduction

The Eurasian badger (*Meles meles*) is a popular animal in the UK that is statutorily protected by the Badger Act 1973, Wildlife and Countryside Act 1981 and Protection of Badgers Act 1992 (Macdonald and Newman 2002) but its existence has been a subject of intense public debate resulting from its role as a wildlife reservoir of some disease-causing organisms including *Mycobacterium bovis* which is the causative agent of bovine tuberculosis (Btb) (Corner et al. 2011). Although, evidence suggests that widespread badger culling in the UK reduced the incidence of tuberculosis in cattle within the culled areas there was an increase in the incidence rate of tuberculosis in adjoining areas due to badger movement (Pope et al. 2007). This highlights the role of animal reservoir component as a facilitator for an increasing transmission rate of bTB and also suggests the need to detect if badgers can play host to some other parasitic diseases.

One of such parasites that have been detected in the UK badgers is trypanosomes. A variety of trypanosome species have been detected so far in British fauna with most being from the subgenera *Herpetosoma* and *Megatrypanum* (Lizundia et al. 2011). Studies conducted on British badgers in Cornwall (Pearce and Neal 1973) and Wytham Woods, Oxfordshire (Lizundia et al. 2011, Macdonald and Newman 2002) have shown them to be infected with *Trypanosoma (Megatrypanum) pestanai* (Fig. 4.1)

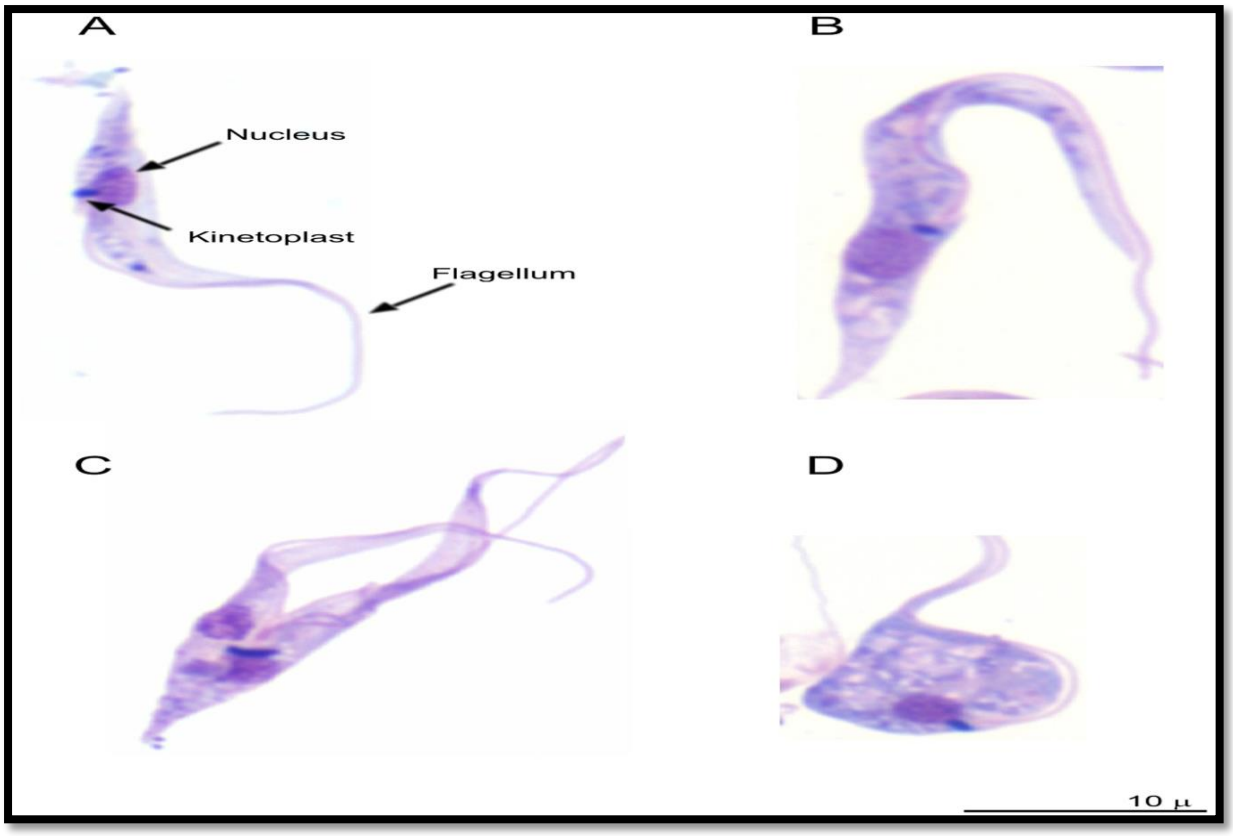


Figure 4.1. Different *T. pestanai* forms in axenic culture following Giemsa stain**.** (A) Epimastigote-like long slender form. (B) Epimastigote-like swollen form. (C) Dividing epimastigote. (D) Large pear-shaped form ("degenerative" form) (Lizundia et al. 2011).

According to Hoare (1972), the non-pathogenic *T. pestanai* was described by Portuguese and French observers to possess a well-developed undulating membrane, free flagellum (4.8-12.0µm long), ovoid kinetoplast near the median nucleus and far from the pointed posterior end, with an overall length ranging from 30-42.0 µm. However, Pearce and Neal (1974) recorded overall length value of 25.6-41.4 µm whereas Lizundia et al. (2011) recorded values of 25.33-33.33 µm. This suggests that there is not yet a definitive value for the overall length of the parasite. Available literature for the parasite still sparse.

Nonetheless, the badger flea (*Paraceras melis*) was identified as the arthropod vector responsible for the transmission of *T. pestanai* to badgers (Lizundia et al. 2011) but

knowledge about infection of British badgers with trypanosomes in other geographical areas is sparse. To date, the only study using molecular tools for the determination of prevalence in infection has been the Wytham Woods study that used an 18S gene PCR to determine the prevalence of *T. pestanai* as 31% in their collection of badgers (Lizundia et al. 2011). Although, there are many molecular methods useful for the detection of trypanosomes (Hide and Tait 2009) in humans and known hosts, wildlife hosts may be more problematic due to the presence of potentially unknown trypanosome species. One method that can be used generically on a wide range of trypanosome species is the Internal Transcribed Spacer (ITS) Nested PCR based on generic trypanosome primers (Ahmed et al. 2011, Cox et al. 2005, Cox et al. 2010). The technique is capable of distinguishing between various species of trypanosomes by amplification of the variable length Internal Transcribed Spacer region of ribosomal RNA and producing unique band sizes for each species/subspecies of trypanosome. The expected band size produced using the ITS-Nested PCR has been determined for clinically important trypanosomes (Cox et al. 2005) but not yet for less studied species such as trypanosomes of wildlife. The lack of the DNA sequence for the full ITS-rRNA in the databases for many wildlife trypanosome species limits *in silico* approaches to determining ITS-Spacer size. On the other hand, 18S rRNA sequence data is available for most trypanosome species and has been used for phylogenetic analysis (Haag et al. 1998, Hamilton et al. 2004, Maslov et al. 1996, Maslov and Simpson 1995). Furthermore, studies have reconstructed phylogenies of trypanosomes using concatenation of 18S and GAPDH sequences (Acosta Ida et al. 2013, Cottontail et al. 2014, Viola et al. 2009) but little is known about the phylogenies of trypanosomes generated using 28S rRNA genes. This is partly due to the lack of availability of many trypanosome 28S rRNA sequences. The 28S rRNA for the badger trypanosome, *Trypanosoma pestanai,* is one such absent member.

The primary objective of this study was to investigate the prevalence of trypanosome infections in British badger blood samples from Woodchester Park in Gloucestershire using the ITS Nested-PCR. This would enable us to evaluate the use of this approach in *T. pestanai* in an example wild animal species and establish the diagnostic band size required for *T. pestanai*. Where detected, the 18S r RNA PCR products would be sequenced to confirm the identity of species. The final objective is to derive the 28S rRNA sequence from *T. pestanai* and apply molecular phylogenetic analyses using this marker to compare with the 18S phylogenies.

3.2 Materials and methods

Eighty-two samples of the Eurasian badger blood were collected taken from badgers at Woodchester Park, Gloucestershire, UK as part of a long-term study e.g. (Rogers et al. 1998). Samples were kindly provided by Alexandra Tomlinson and Prof Dez Delahay (Table 4.1).

Table 4.1 Badger sample codes

	Badger sample codes					Total
	24P	9P	13F	32N	82W	
	59B	28K	34P	45R	28F	
	10N	2P	12K	14L	41L	
	13L	15K	58R	24L	11K	
	26K	38K	7K	67Y	45P	
	3K	50N	38L	21K	8I	
	1K	31K	13N	40K	15L	
	17K	25N	66B	5L	13K	
	39K	1L	9N	33K	26N	
	22K	40L	9L	23N	25K	
	17F	31L	33Y	2K	63Y	
	2F	4K	30L	24K	9K	
	36K	14N	37K	29K	35P	
	38Y	34K	32K	29N	17L	
	5K	27K	23K	30K	8L	
	12L	35K	41K	6L	10K	
	2N	22N				
Total	**17**	**17**	**16**	**16**	**16**	**82**

Data was also collected on each badger. This included gender, the sett where the badger was raised and where the badger now resides. Also, for each badger, the scaled mass index (SMI) - a modification of the condition index (Peig and Green 2009) was calculated. Badgers with an SMI of <1 indicated a poorer condition while SMI of >1 indicated a better condition. For the purposes of analysis of evenness of trypanosome infection tested in this study, the badger population was divided into North and South using a line of demarcation which was drawn separated by the Woodland Park ponds (Brick Kiln pond, Middle pond and Kennel pond) (Fig. 4.2). The blood sample from each badger was stored at -20 ^{0} C before being processed in the laboratory as follows;

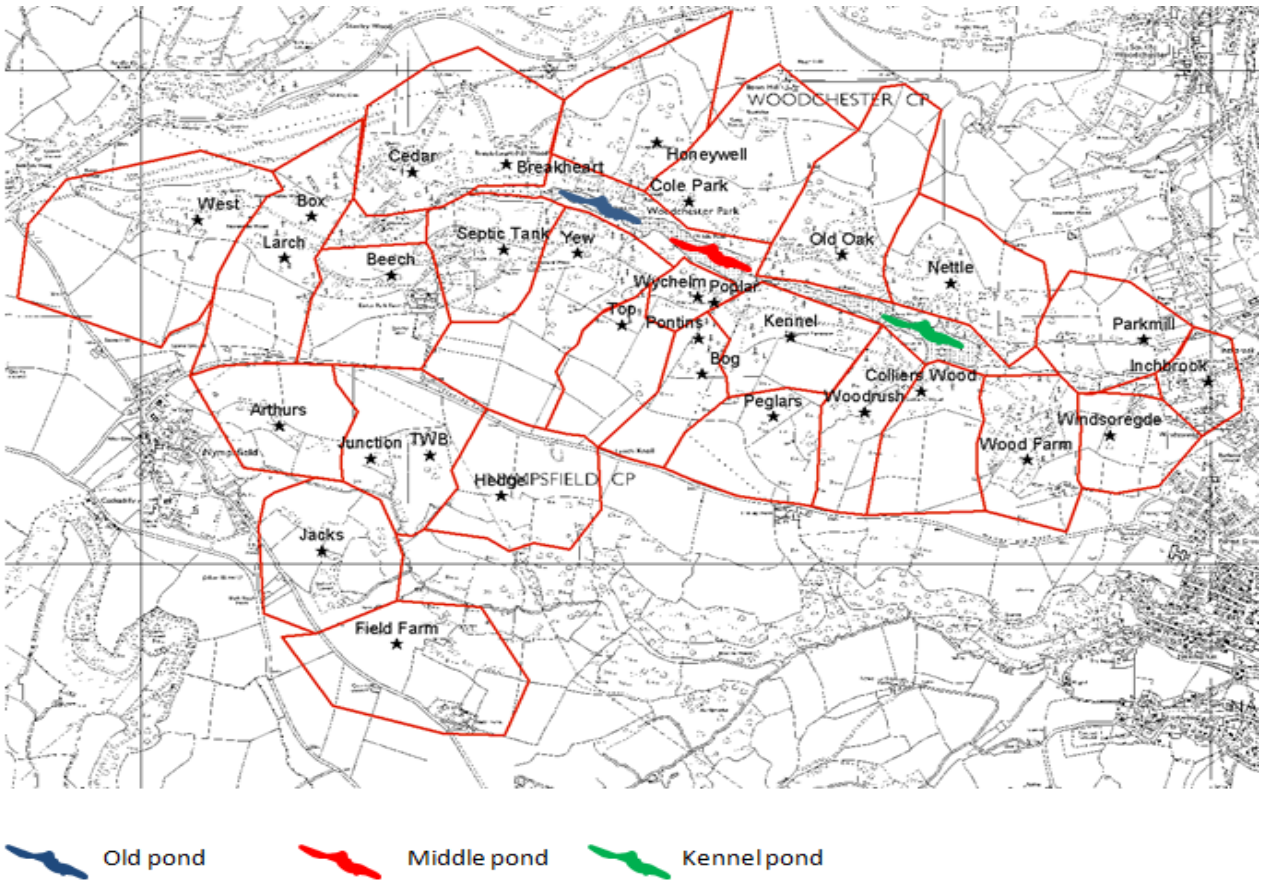


Figure 4.2 Map showing badger setts. A line of demarcation was drawn separated by the ponds grouping badgers into north or south based on their sett location on the map.

Full details of protocols can be found in Chapter 2 but briefly, the badger blood samples were extracted using a modification of a phenol-chloroform protocol described previously (Morley et al. 2005, Morley et al. 2008) but with appropriate measures to prevent contamination (Bajnok et al. 2014, Williams et al. 2005). The tubulin PCR was applied to test extracted DNA for ability to amplify and to confirm good quality of DNA (Duncanson et al. 2001, Terry et al. 2001). Following this, extracted DNA from the badger blood was tested for trypanosome infection using the ITS-Specific (nested) PCR method (Cox et al. 2005). Primers were then designed (see Chapter 2) and a novel 18S PCR was applied for a further identification of species of trypanosomes detected. A new 28S PCR was developed and applied for the derivation of a novel 28S rRNA gene for the species of trypanosomes detected. PCR products after amplification were sourced out to a sequencing company (Source Bioscience) for purification and sequencing to confirm the identity of trypanosomes using bioinformatics. Statistical analysis was carried out using IBM SPSS Statistics for Windows (version 20), and lastly molecular phylogenetics reconstructed using trypanosomes 18S sequences and concatenated sequences (18S and 28S sequences) was applied to determine the taxonomic status of the badger trypanosomes. All protocols used for this study are provided in detail in Chapter 2.

4.3 Results

4.3.1 Extraction of DNA from badger blood samples

The aim of this objective was to successfully extract DNA from all 82 badger blood samples collected from Woodchester Park to investigate them for infection with trypanosomes. Using phenol-chloroform DNA extraction procedure described above (and Chapter 2.2.1), the DNA from 82 badger blood samples was extracted. However, prior to testing whether the samples could be amplified using PCR, two samples of the newly extracted badger DNA (37K and

59B) were tested for presence of DNA by mixing 2 µl of the badger DNA with 2 µl of loading buffer and running the mixture on a 1% agarose gel at 70 V. Results from the gel indicated the extraction was successful as there were clear bands which represent the presence of DNA in the samples (Fig. 4.3).

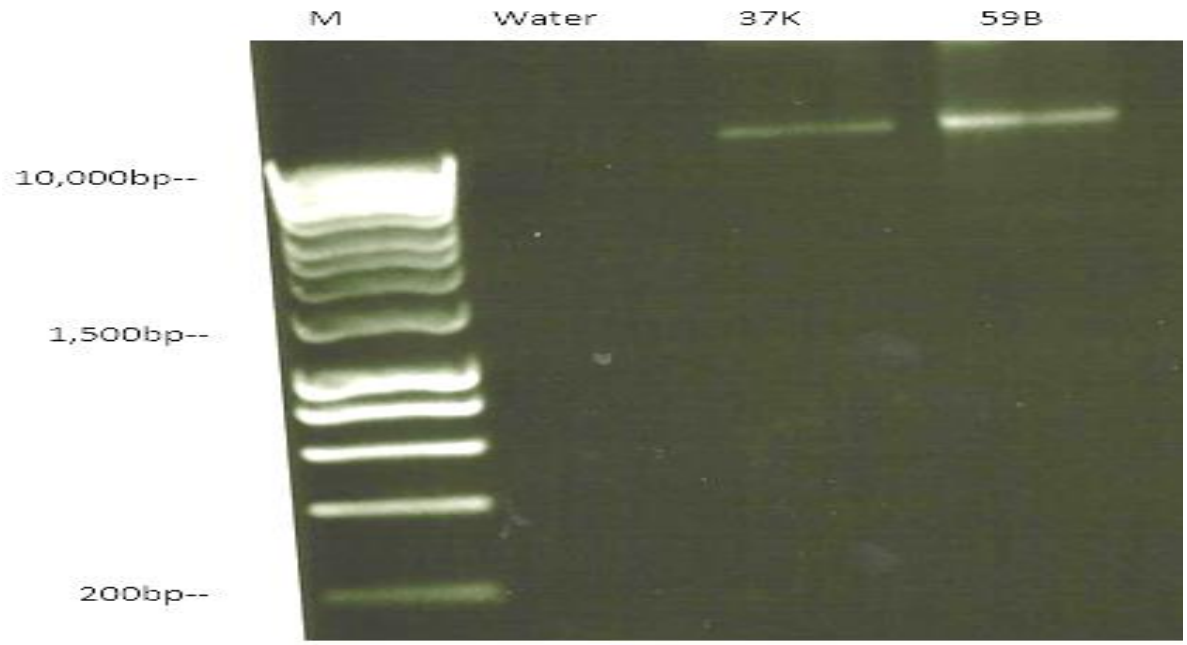


Figure 4.3. Gel electrophoresis of DNA extracted from the badger blood. DNA seen as white bands in 37K and 59B indicating a successful extraction of DNA from the badger blood samples. Water was used as a negative control. M represents 1 HyperLadder 1Kb plus marker which shows DNA fragments of known sizes.

Following this, the newly extracted DNA from 4 other badger samples (24L, 17K, 45P and 45R) were also tested for the presence of DNA using the same procedure described above. This was necessary to investigate if this standard method of DNA extraction could be applied to the badger blood. Interestingly, the results from the gel indicated the extraction was successful based on visualization of clear bands on the gel (Fig. 4.4). The DNA from the remaining badger samples to be used for the study were then extracted using the phenol-chloroform procedure as detailed in Chapter 2.2.1

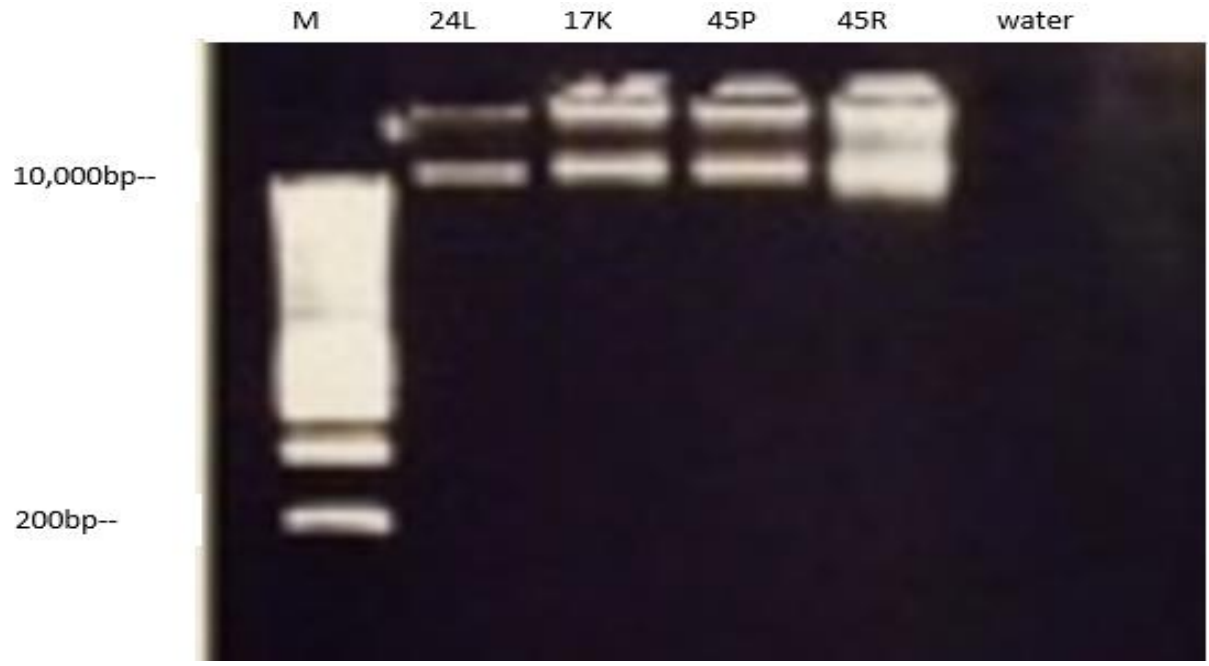


Figure 4.4. Gel electrophoresis of DNA extracted from some more badger blood. DNA seen as white bands in 24L, 17K, 45P and 45R lanes indicating a successful extraction of DNA from the badger blood samples. Water was used as a negative control. M represents 1 HyperLadder 1Kb plus marker which shows DNA fragments of known sizes. The smear observed in lane M was due to too much volume (10 ul) of the marker loaded instead of 5ul.

4.3.2 PCR based quality test for DNA amplification

The aim of this objective was to determine if the newly extracted DNA can be amplified using PCR. This was achieved by PCR amplification using generic primers designed against the mammalian tubulin genes. Successful amplification using these genes is indicative of good quality DNA that can be amplified by other primers. The DNA from two badger samples (37K and 59B) successfully extracted using phenol-chloroform, as described above, were used to test this technique. Detailed protocol is available in chapter 2.3. After PCR amplification, 10 µl of the PCR products from each of the two samples were mixed in separate tubes with 5 µl of loading dye and loaded into a 1.0 % agarose gel run at 100 V. Finally, the gel was visualised under UV light which showed that the extraction produced good quality DNA as there were clear bands present in the lanes on the gel indicating that the badger DNA can be amplified using PCR (Fig. 4.5).

Figure 4.5. PCR amplification of badger DNA using generic tubulin primers. The bands on badger samples 37K and 59B indicate their DNA can be amplified using PCR. M represents HyperLadder 1Kb plus DNA marker while water was used as a negative control.

Following the successful exploration of this technique, all badger samples used in this study that tested negative for ITS-Nested PCR were amplified using this method to ensure good quality of DNA was extracted from the negative samples (See Table 4.2).

4.3.3 Development of the ITS-PCR for detection of trypanosomes from known infected DNA samples

To determine the capability of a set of ITS nested primers to amplify the target region of 2 known trypanosome infected samples, two samples were provided by Prof. Geoff Hide (Ph.D. Supervisor), one (coded 2848) was infected with *T. b. brucei* while the other (coded 2856) was infected with an unknown species of trypanosomes. The ITS-Nested PCR produces unique band sizes for different species/subspecies of trypanosomes and those for clinically important African trypanosomes have been specified. For *T. brucei* the expected band size is 1,216 bp (Cox et al. 2005) but the band size of the unknown species (2856) to be

used also as control was unknown. Successful amplification of *Trypanosoma* from these samples using the ITS-Nested primers would imply they can be used as control samples for subsequent ITS-Nested PCR amplification. The following set of primers are used for the double round ITS-Nested PCR; First round PCR, uses outer primers, ITS 1 (5'-GAT TAC GTC CCT GCC ATT TG-3') and ITS 2 (5' TTG TTC GCT ATC GGT CTT CC-3') while second round PCR uses the inner primers- ITS 3 (5' GGA AGC AAA AGT CGT AAC AAG G- 3') and ITS 4 (5' TGT TTT CTT TTC CTC CGC TG-3') (Cox et al., 2005). Amplification was achieved by using the protocol described in Chapter 2.4 above. Following visualisation under UV light, the result showed clear bands from both samples (2848 and 2856) indicating a successful amplification of the ITS primers to the ITS region and thus indicating the presence of trypanosome in the known trypanosome infected samples (Fig. 4.6). As expected, sample 2848 (*T. brucei*) produced a band size of about 1216bp while sample 2856 produced a band size of about 850bp.

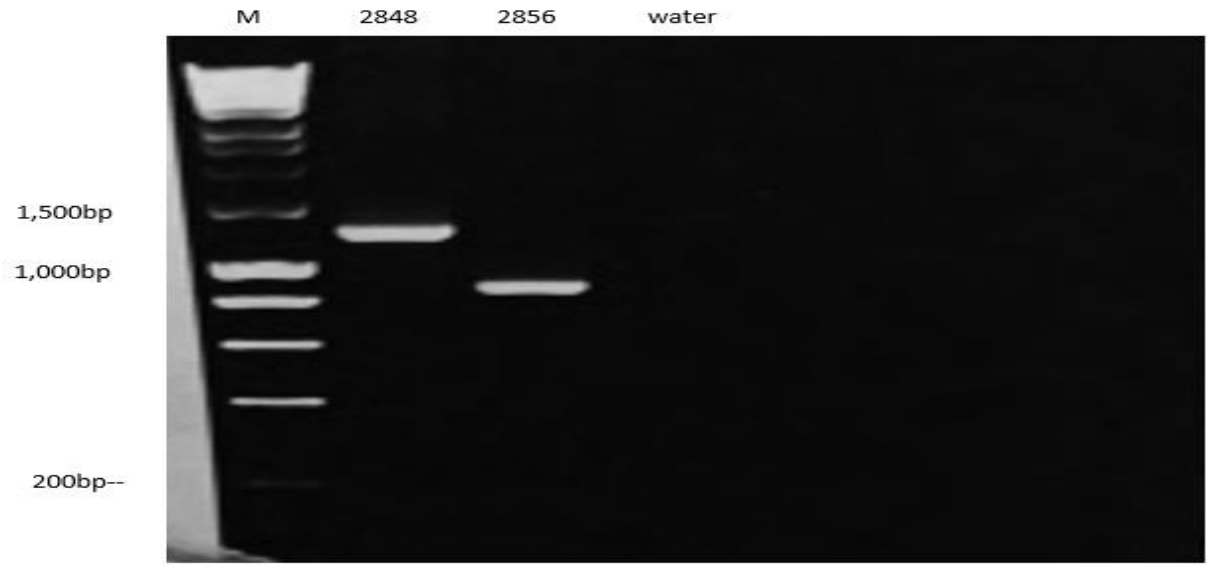


Figure 4.6. ITS-Nested PCR amplification of trypanosome control samples. 2848 represents *T. brucei* (~1216bp) while 2856 represents unknown species (~850bp). M represents HyperLadder I Kb Plus DNA marker while water was used as the control sample.

4.3.4 ITS-Nested PCR: Testing the badger samples

This objective was targeted at determining if the badger DNA samples would amplify with ITS primers and thus identify the primers were detecting trypanosomes in the samples. This was achieved using the ITS-Specific (Nested) PCR described above (Chapter 2.4). Twenty-nine out of 82 badger DNA amplified using this approach produced band sizes of about 1271bp indicating the badgers are positive for trypanosome, resulting in a 35.4% (25.9% - 46.2%; 95% CI) prevalence (Table 4.2; Fig. 4.7).

Table 4.2. Badger sample codes vs trypanosome infection and tubulin positive

Badger ID	Trypanosome +ve	Tubulin +ve	Badger ID	Trypanosome +ve	Tubulin +ve	Badger ID	Trypanosome +ve	Tubulin +ve	Badger ID	Trypanosome +ve	Tubulin +ve
24P	1	0	2P	0	1	12K	1	0	14L	0	1
59B	0	1	15K	0	1	58R	1	0	24L	0	1
10N	0	1	38K	0	1	7K	0	1	67Y	0	1
13L	1	0	50N	1	0	38L	1	0	21K	1	0
26K	1	0	31K	1	0	13N	0	1	40K	0	1
3K	0	1	25N	1	0	66B	0	1	5L	1	0
1K	0	1	1L	0	1	9N	1	0	33K	1	0
17K	0	1	40L	0	1	9L	0	1	23N	0	1
39K	0	1	31L	0	1	33Y	0	1	2K	0	1
22K	0	1	4K	1	0	30L	0	1	24K	0	1
17F	1	0	14N	1	0	37K	0	1	29K	1	0
2F	0	1	34K	0	1	32K	0	1	29N	1	0
36K	0	1	27K	0	1	23K	0	1	30K	0	1
38Y	1	0	35K	1	0	41K	0	1	6L	0	1
5K	0	1	17L	1	0	8L	0	1	10K	0	1
12L	1	0	28F	0	1	41L	1	0	11K	0	1
35P	0	1	8I	0	1	15L	1	0	13K	0	1
82W	0	1	25K	0	1	63Y	1	0	9K	0	1
45P	1	0	13F	1	0	32N	0	1	3N	0	1
26N	0	1	34P	0	1	45R	0	1	22N	0	1
9P	1	0	28K	1	0						

** 1 in table indicates “yes” throughout table while 0 indicates “no” after PCR. Trypanosome +ve status was determined by ITS-Nested PCR while tubulin +ve was determined using the DNA quality test PCR.

As shown in Table 4.2 above, badger samples that tested negative for trypanosomes using the ITS-nested PCR were amplified using the DNA quality test PCR, to ensure that quality DNA capable of amplification by PCR has been isolated via the phenol-chloroform extraction method.

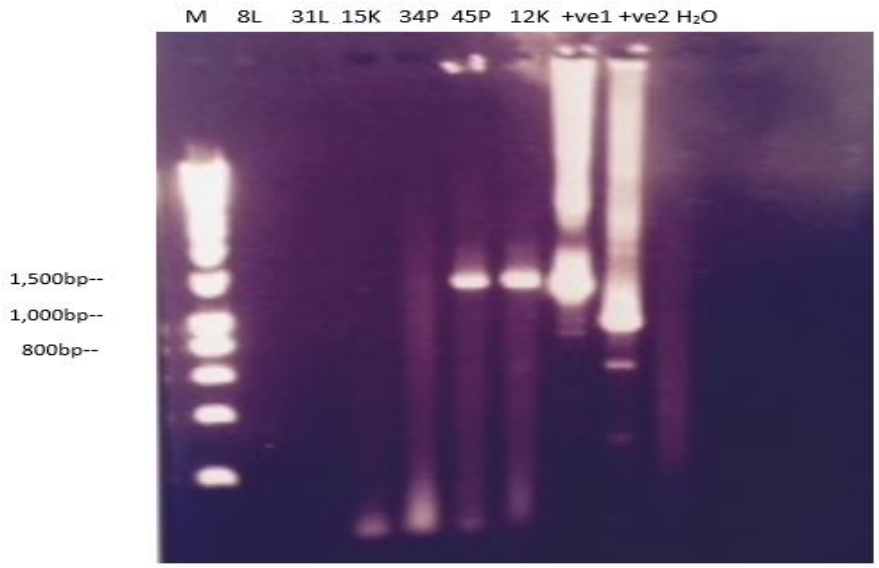


Figure 4.7. ITS-Nested PCR amplification of badger samples. Lanes 8L, 31L, 15K, 34P 45P and 12K represents badger samples. Bands shown on 45P and 12K (~1271bp) indicates positive for trypanosomes while no bands in the other samples (8L, 31L and 15K) indicate negative for trypanosomes. +ve1 is 2848 control sample (*T. brucei*, ~1216bp) while +ve2 is 2856 control sample (unknown species, ~850bp). M represents HyperLadder I Kb Plus DNA marker while water was used as a negative control sample.

As shown in figure 4.7, there were unspecific bands or smears appearing in lanes for control samples. Although, the DNA concentration of the control sample was not checked, a serial dilution of control sample 2848 and further amplification of this sample using the ITS-Nested PCR was carried out to confirm if it were rather due to high concentration of DNA in control samples. Interestingly, the results shown after gel electrophoresis indicated it was indeed due to a high concentration of DNA. The bands became clearer with more dilutions until the optimal concentration required for ITS-Nested PCR amplification was exceeded (Fig. 4.8).

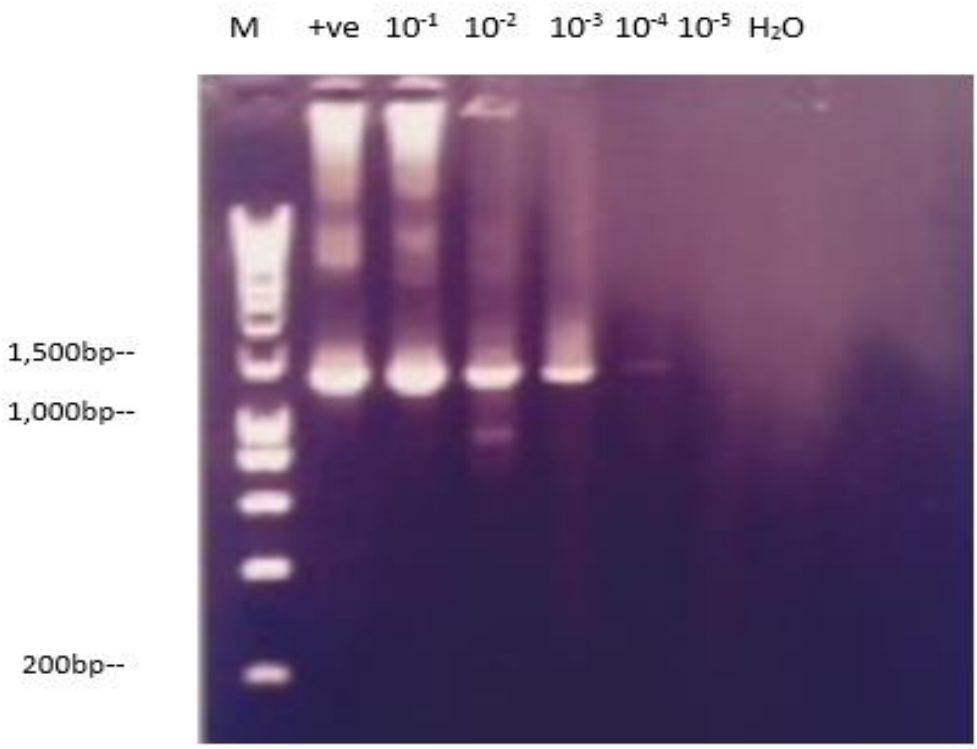


Figure 4.8. Serial dilution of 2848 (*T. brucei* control sample). +ve is the stock 2848 sample, 10^{-1} - 10^{-5} are the dilutions carried out from the original stock. As shown on the gel the bands became clearer following series of dilution and at 10^{-3} of the dilution factor the amplification yielded a specific and strong band (~1216bp) for *T. brucei*. M represents HyperLadder I Kb Plus DNA marker while water was used as a negative control sample.

Having established the smears were not due to contamination, 2848 was used throughout as the trypanosome control sample during the ITS-Nested PCR amplification of all badger samples.

It can be recalled that (Cox et al. 2005) described the ITS-Nested PCR method as capable of detecting the inter-specific length variation of the ITS regions of ribosomal genes producing a unique size of PCR product for each species of trypanosomes (Table 4.2).

All samples positive for trypanosomes were of similar band sizes (~1271bp) implying they might be infected with the same species of trypanosomes. Although, the nested PCR detected trypanosomes in the British badgers further analysis of their sequence was required to ascertain what species of trypanosomes have been detected in the badgers.

Therefore, PCR products along with their associated primers (ITS3 and ITS4) were sent to Source Bioscience for purification and sequencing. However, the sequencing did not produce quality sequence data using these primers after several attempts (Fig. 4.9).

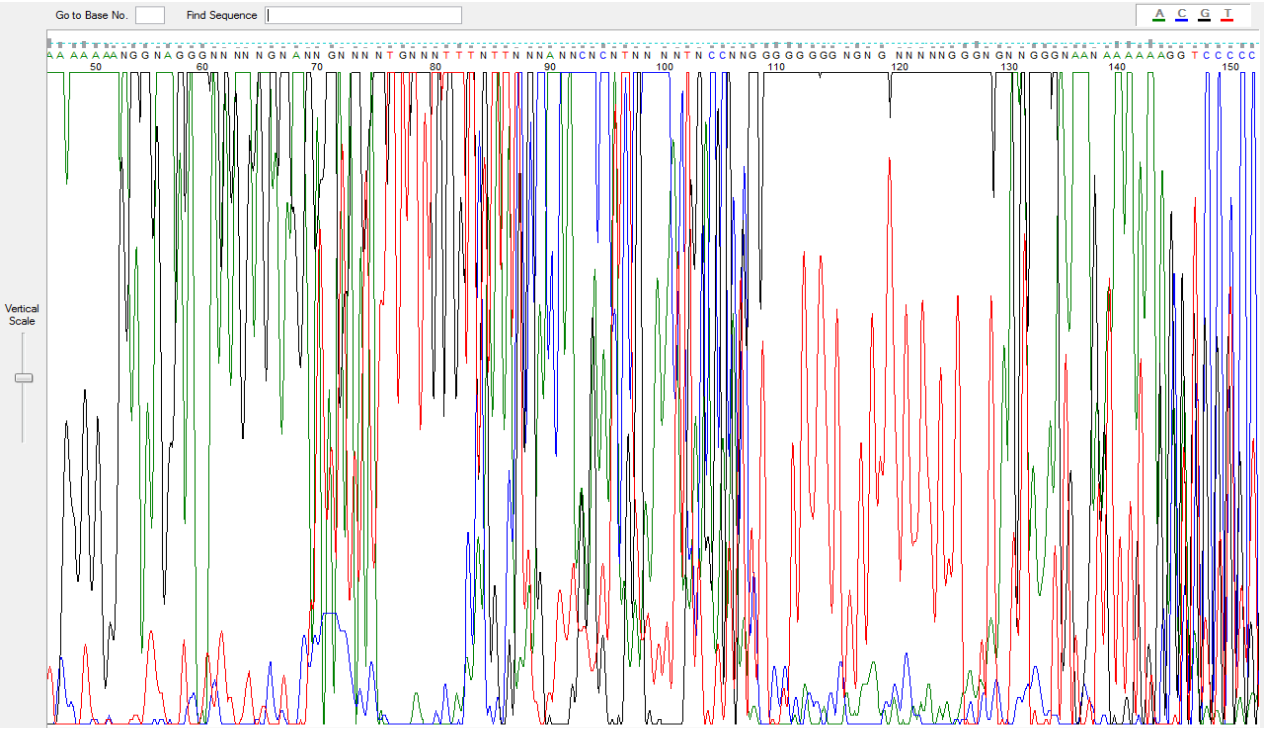


Figure 4.9. Chromatogram of a poor quality badger sequence data. The Finch TV was the software used for analysis. The nucleotide bases are called as peaks and annotated by colours (Green for Adenine, blue for cytosine, black for guanine and red for tyrosine). The abundance of N's (miscalled nucleotides) in the chromatogram indicates poor DNA quality.

Based on this, it was impossible to confirm the identity of the badger trypanosomes by analysing the derived sequences. A new set of primers (ITS5 and ITS6) targeting the inner part of trypanosomes 18S and 28S rRNA gene (Fig. 4.10) was designed for a third reaction

using the same reaction conditions as the previous to amplify the PCR products derived from the previous reaction with ITS3 and ITS4. Details of ITS5 and 6 described in Chapter 2.4.2.

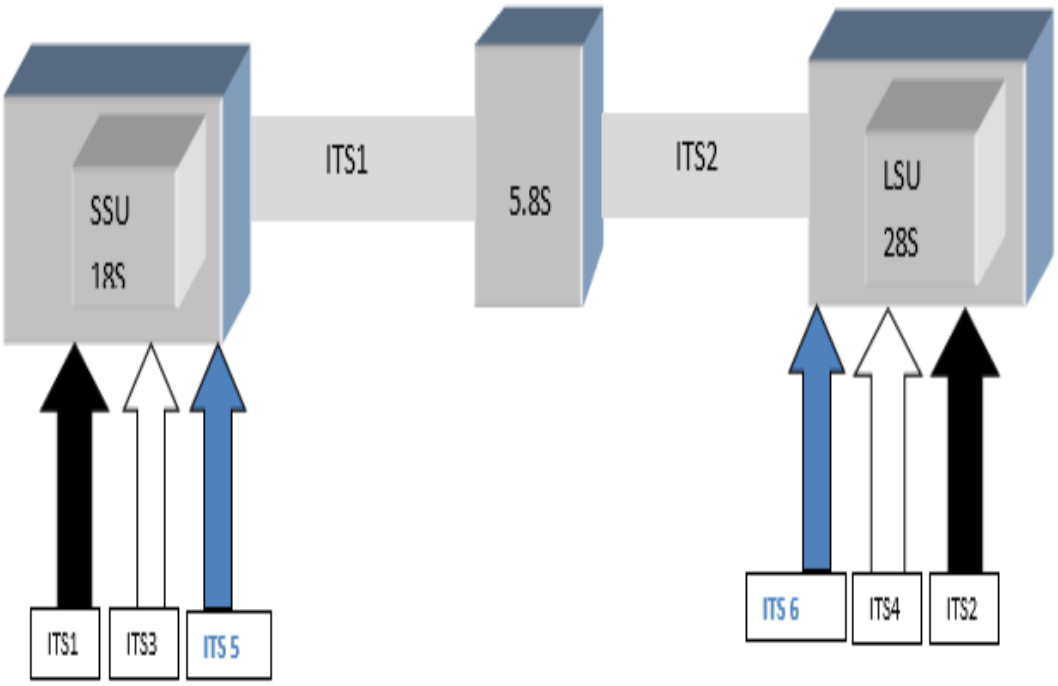


Figure 4.10. Schematic representation of regions of the rRNA gene locus targeted for nested PCR. The conserved regions are the small subunit (SSU), 5.8S and large subunit (LSU). The two spacers (ITS1 and ITS2) vary in size between species and occasionally subspecies of trypanosomes. The nested primers are represented by black arrows (outer primers) ITS1 and ITS2 and white arrows (inner primers) ITS3 and ITS4 while blue arrows represent the 3rd reaction primers ITS 5 and ITS 6.

This was necessary to confirm that the initial sets of primers (ITS1, ITS2, ITS3 and ITS4) were detecting trypanosomes and to try and derive sequence data using the new set of primers. Amplification of the PCR products from trypanosome positive samples (45P, 12K and 2848) using ITS 5 and ITS 6 produced clear bands which again suggested the badgers were infected with trypanosomes (Fig. 4.11).

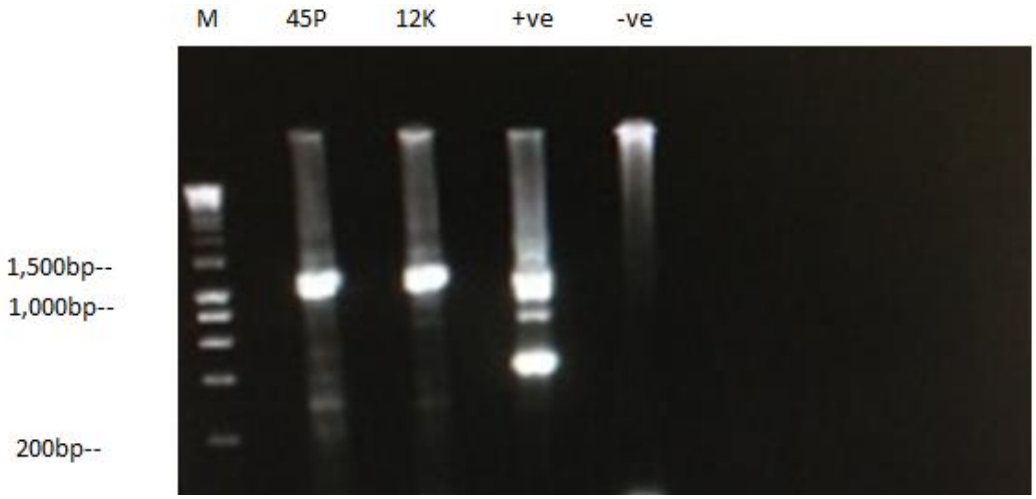


Figure 4.11. ITS-Nested PCR using ITS5 and ITS6. Bands on trypanosome positive samples (45P, 12K, +ve/2848) amplified using new internal primers ITS5 and ITS6 suggest the samples are infected with trypanosomes. Lane M is the HyperLadder 1kb marker. –ve is water used as a control sample.

Sequencing of the PCR products again failed to yield good sequence data, so it was impossible to ascertain what species of trypanosomes was detected. Hence, it was hypothesised that the difficulty in deriving good sequences from the internal primers (ITS3, ITS4, ITS5 and ITS6) might be due to the ITS region for most trypanosomes being variable and not conserved and/or problematic to be derived using normal Sanger sequencing.

So the next objective was to derive sequence data with the outer primers (ITS1 and ITS2) since they would contain a greater proportion of the conserved regions of 18S and 28S rRNA genes of trypanosomes on successful amplification. To achieve this, the badger samples were amplified with ITS1 and ITS2 and the PCR products were re-amplified in a new reaction with the same set of primers (ITS1 and ITS2) using the same ITS-Nested reaction condition (see Chapter 2.4). This produced very clear bands suggesting that the badgers were infected with trypanosomes (Fig. 4.12).

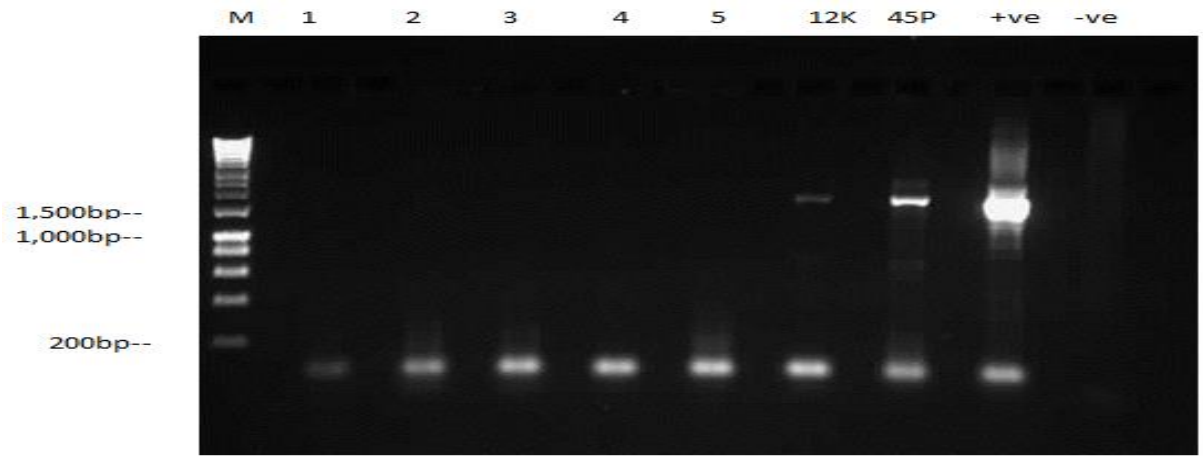


Figure 4.12. First round ITS-Nested PCR. Gel image showing size of the band from badger samples amplified using primers ITS1 and ITS2. 12K and 45P are badger samples positive for trypanosomes. Lane M is the Hyperladder 1 marker. 1-5 are cattle samples for a different study (Chapter 3). +ve is 2848 and –ve are control samples.

A linear regression graph was plotted using Microsoft Excel package that generated an equation which we used to deduce an estimate of the band size on the gel image, our fragment migrated to 2.4cm down the gel so from the equation generated (y = _0.4989x + 4.417) the size of the band obtained was estimated at 1652bp (Fig. 4.13).

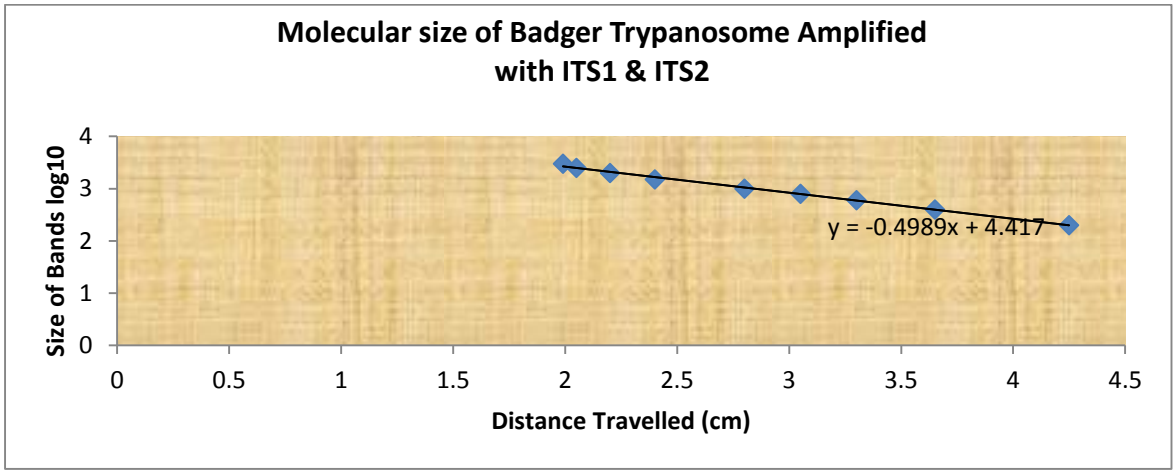


Figure 4.13. Linear Regression graph plotted to estimate the size of the band from badger samples. From the equation "x" (2.4cm) represents the distance migrated by the PCR fragment from the arbitrary line. Y- axis represents the log of fragment of known sizes (bp) used as marker (Hyperladder1) while X- axis represents the distance the fragments migrated from the chosen arbitrary line (cm) from the gel image.

However, a further sequencing of PCR products produced DNA sequences which after analyses managed to produce a partial alignment with *Trypanosoma (Herpetosoma) otospermophili* (Table 4.3; Appendix A). *T. otospermophili* is a parasite of ground squirrels and not a badger trypanosome, belonging to *Herpetosoma* subgenus with host specificity a major criterion for classifying trypanosomes in this subgenus (Haore 1972).

Table 4.3 BLAST hits showing alignment with ITS region of *T. otospermophili*.

TITLE	BLAST HIT RESULT	E-Value	Max Identity	Alignment Region
12K ITS1	*Trypanosoma otospermophili* genes for IGS, 18S rRNA, 5.8S rRNA, 28S rRNA and ITS1-6 (AB175626.1)	1e-48	97%	2134-2194bp
12K ITS2	*Trypanosoma otospermophili* genes for IGS, 18S rRNA, 5.8S rRNA, 28S rRNA and ITS1-6 (AB175626.1)	4e-133	95%	3621-3321bp
45P ITS1	*Trypanosoma otospermophili* genes for IGS, 18S rRNA, 5.8S rRNA, 28S rRNA and ITS1-6 (AB175626.1)	4e-82	95%	2648-2828bp
45P ITS2	*Trypanosoma otospermophili* genes for IGS, 18S rRNA, 5.8S rRNA, 28S rRNA and ITS1-6 (AB175626.1)	8e-40	97%	3413-3353bp

**Title represents badger samples and primers used for sequencing the PCR product. **Blast hit result represents DNA sequences homologous to the badger sequences. **Expect (E)-value is the number of hit expected to see by chance, the lower the E-value the more significant the result. **Max Identity represents percentage of similarity between sequences. **Alignment region represents the area of NCBI published sequence covered by the DNA sequence entered into the software program.

To confirm if the parasite was actually *T. otospermophili,i* or something different, the 18S PCR was chosen as the preferred technique. This is because of the abundant availability of 18S sequences of most trypanosomes on the databases which would allow for easy comparison of the similarity of the 18S sequence of the badger trypanosome with the 18S

sequence of other trypanosomes so as to ascertain the species of trypanosomes detected in the badgers.

4.3.5 Amplification of the small subunit (SSU): 18S PCR

Primers targeting the 18S region of trypanosomes rRNA were designed and the 18S PCR was carried out on samples that indicated positive for trypanosomes, for DNA sequencing purposes to identify the species of trypanosome detected and confirm that the ITS-PCR technique was indeed detecting the correct species in the British badgers. (See Chapter 2.5 for protocols). The result produced clear bands indicating a successful amplification and detection of trypanosomes in the badger samples. The 18S PCR produced a band size of 1864bp (Fig. 4.14) which was the expected band size for *T. pestanai* deduced using the 18S primers designed for this study but DNA sequencing was still needed to confirm identity via bioinformatics.

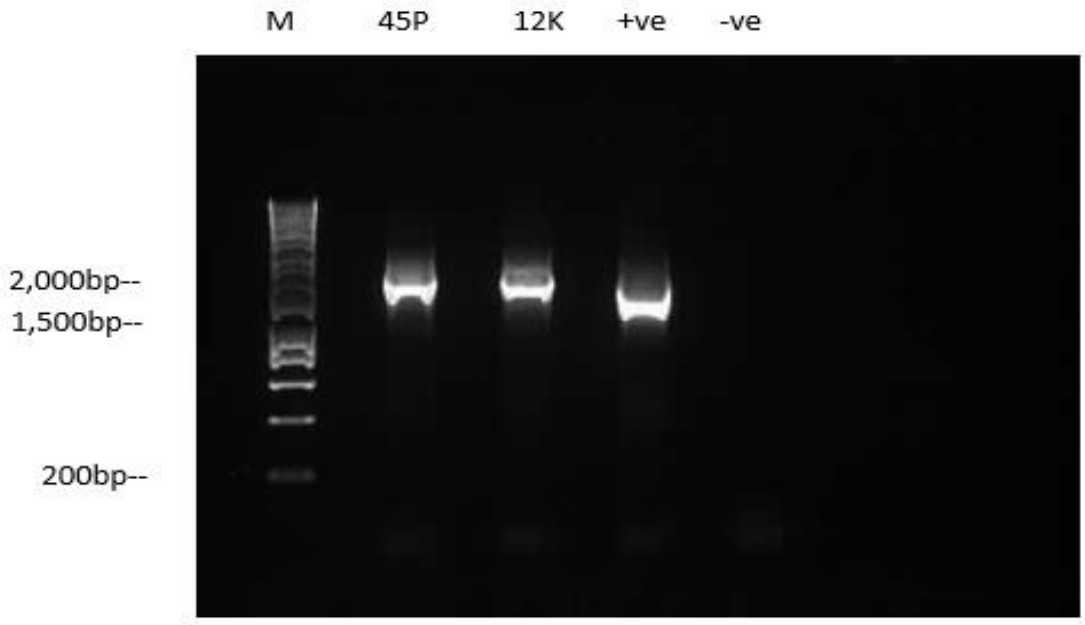


Figure 4.14. Products of 18S PCR amplification. Badger DNA (45P and 12K) successfully amplified with 18S primers, producing band sizes of 1864bp which indicates infection with *T. pestanai*. +ve represents 2848 (*T. brucei*) control sample (1815bp). M represents HyperLadder 1kb marker. –ve is water used as control.

The PCR products containing bands of interest must be purified prior to Sanger sequencing in order to produce interpretable sequence result during analysis. Hence, both the purification and sequencing of the PCR products were achieved by following the procedure described in Chapter 2.12. Analysis of the sequence data was performed using the program Geospiza Finch TV (http://www.geospiza.com) which showed clear evenly spaced nucleotide peaks easy to call with no baseline noise (Fig. 4.15).

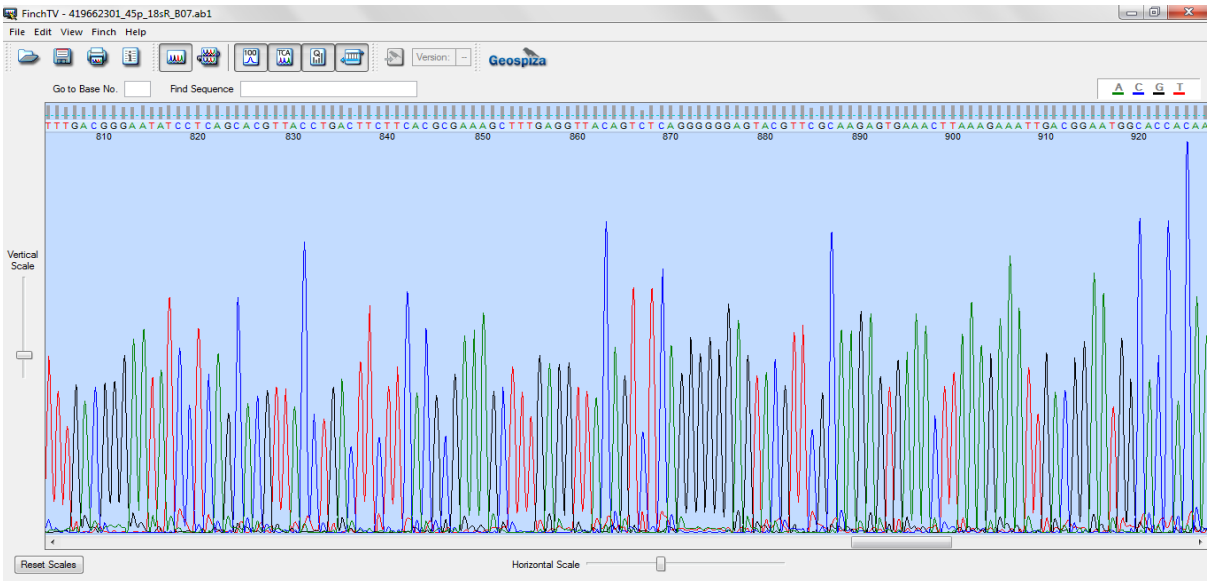


Figure 4.15. Chromatogram of *T. pestanai* DNA sequence data. The nucleotide bases are called as peaks and annotated by colours (Green for Adenine, blue for cytosine, black for guanine and red for tyrosine). The clear peaks in the chromatogram indicate good DNA sequence quality.

The DNA sequence was retrieved from the Finch TV programme on FASTA format and used to search for "highly" identical matches of the sequence on the NCBI database (http://www.ncbi.nlm.nih.gov/) using the Megablast program which produced BLAST hits confirming a 100% alignment with the ITS region of *Trypanosoma (Megatrypanum) pestanai* gene with E-values at zero (Appendix B). The lower the E-value or the closer it is to zero, the more significant the match is (Zhang et al. 2000). Thus, our samples were statistically shown to be *T. pestanai*. The difficulty in obtaining alignment with the 28S of *T. pestanai* rRNA gene was based on the unavailability of this region of the parasite genome on the NCBI

database. The 28S rRNA gene could also play an important marker for constructing phylogenies of trypanosomes, but little is known about the phylogenies of trypanosomes generated using 28S rRNA genes. This may be partly due to the lack of availability of many trypanosome 28S rRNA sequences. The 28S rRNA for the badger *Trypanosoma pestanai* is one such absent member.

4.3.6 Amplification of the large subunit (LSU): 28S PCR

The 28S ribosomal rRNA of trypanosomes was the target region for PCR amplification. The 28S sequence for *T. pestanai* was not available in the databases. Suitable primers for novel amplification of the 28S rRNA gene from *T. pestanai* were designed by assembling 28S sequences of trypanosomes; *T. otospermophili* (GI: 46091661), *T. kuseli* (GI: 46091662), *T. rangeli* (GI: 662247341) and *T. minasense* (GI: 159157536), accessible from NCBI and the 28S PCR was carried out on badger samples that previously tested positive for trypanosomes (See protocols in Chapter 2.6). Initial attempts to amplify the 28S region of trypanosomes RNA from positive samples were problematic (Fig. 4.16), however, after several optimisation steps and change of enzyme the 28S region was successfully amplified (Fig. 4.17) producing clear bands of ~2460bp which were sequenced. The 28S rRNA sequence for *T. pestanai* derived from this study is available (Accession Number: KR527480; Appendix C).

Figure 4.16. Gel image showing unsuccessful 28S PCR. The reaction mixture and thermocycling conditions were same as the first round ITS-Nested PCR (Chapter 2.6). Lane M is the Hyperladder 1 marker of known molecular weight, 45P, 12K and 2848 are known trypanosome +ve samples and –ve is water used as a control sample.

Figure 4.17. Gel image showing successful 28S PCR. The reaction mixture was same as the 18S PCR (Chapter 2.6) but the thermocycling conditions were 1 min at 95^0 C, followed by 25 cycles of 15s at 95^0 C, 15s at 54^0 C, 1.30 min at 72^0 C and a final cycle of 5 min at 72^0C. Lane M is the Hyperladder 1 marker of known molecular weight, 45P, 12K and 2848 are known trypanosome +ve samples and –ve is water used as a control sample.

4.3.7 Molecular phylogenetic analysis

To investigate the phylogenetic relationship of *T. pestanai* with other trypanosome species, two set of trees were generated. Firstly, a tree based on existing 18S (SSU-rRNA) trypanosomes sequences was constructed (Fig. 4.18) for comparison with a second concatenated tree (Fig. 4.19) based on 18S and 28S sequences (i.e. SSU-rRNA + LSU-rRNA). Briefly, the DNA sequences accessed from NCBI were aligned Muscle (v3.7), Gblocks software was used to trim sequences. This procedure was performed using Phylogeny.fr webserver (See appendix D for 18S alignments and E for 18S and 28S alignments). MEGA (v6) was used to select model and construct trees. More details are available in Chapter 2.14.

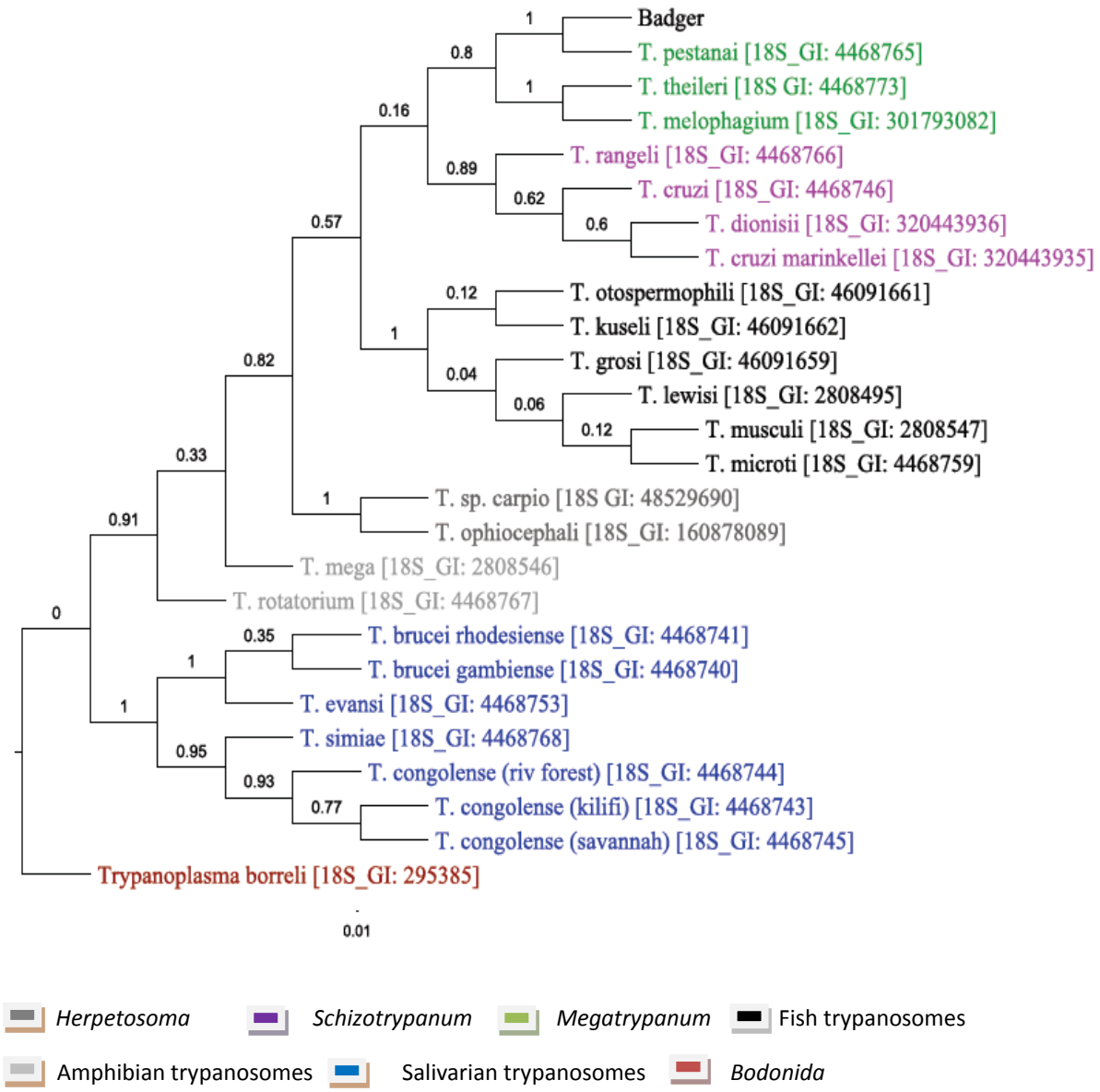


Figure 4.18. Phylogenetic tree of SSU-rRNA sequences of trypanosomes implemented using the Maximum Likelihood method based on the Kimura 2-parameter model (Kimura 1980). The tree with the highest log-likelihood (-2004.6901) is shown. Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.1500)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 26 nucleotide sequences. There was a total of 716 positions in the final dataset. The numbers after species name on branch are GenInfo Identifier number (GI) while annotated colours indicate different groups of kinetoplastids represented in small boxes above. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

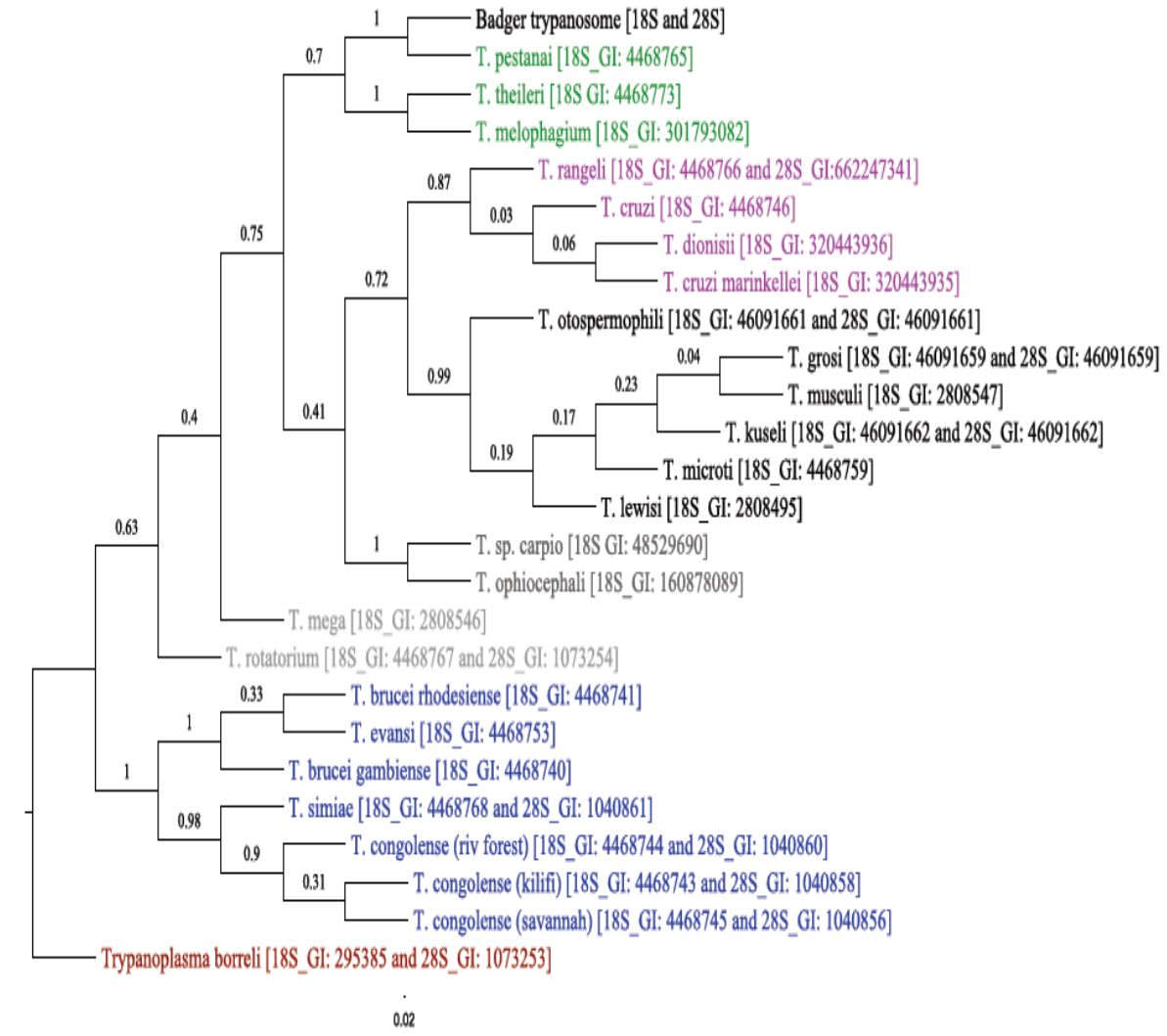


Figure 4.19. Phylogenetic tree of concatenated SSU-rRNA and LSU-rRNA sequences of trypanosomes implemented using the Maximum Likelihood method based on the Kimura 2-parameter model(Kimura 1980). The tree with the highest log likelihood (-4361.5258) is shown. Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.1323)). The analysis involved 26 nucleotide sequences, 11 of which, used the concatenated datasets (*T. grosi, T. otospermophili, T. kuseli, T. rangeli, T.pestanai, T. rotatorium, T. simiae, T. congolense (riverine forest), T. congolense (kilifi), T. congolense (savannah), Trypanoplasma borreli* and badger trypanosome). There were a total of 1560 positions in the final dataset. The numbers after species name on branch are GenInfo Identifier number (GI) while annotated colours indicate different groups of kinetoplastids. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

Trypanosoma Borelli, a member of Bodonidae, was used to form an out-group based on its close relationship with Trypanosomatidae and its clade can easily be distinguished from the genus *Trypanosoma* in both the SSU-rRNA (Fig. 4.18) and Concatenated Tree (Fig. 4.19). The trypanosomes were divided into two major clades, one representing the salivarian trypanosomes while the other represented the stercorarian trypanosomes with the division being obvious in both trees (Fig. 4.18 and Fig. 4.19). The *Megatrypanum* subgenus (Haore 1972), appear to cluster together with a high bootstrap support value of 80% on the SSU-rRNA tree (Fig. 4.18) and 70% on concatenated tree (Fig. 4.19). Analysis of both trees showed that the badger trypanosome sequences, determined from the Woodchester badgers, positioned in the *Megatrypanum* clade and clustered together with *T. pestanai* with a high bootstrap value of 100% on both trees (Fig. 4.18 and Fig. 4.19). This confirmed their identity as *T. pestanai* rather than any closely related species.

4.3.8 Statistical analysis: dynamics of trypanosoma pestanai infection in Woodchester Park

In order to investigate whether the distribution of infected animals (35.4% prevalence) was evenly spread throughout the Woodchester Park sample site, the prevalences of trypanosome infection were examined with statistical analysis (IBM SPSS Statistics for Windows version 20) using gender, body condition and different groups as parameters. Although, more females (55.2%) were infected than males (44.8%), there was no significant association between trypanosome infection and gender ($p = 0.771$, Chi-Square test). Also, the majority of badgers (86.6%) were in poor condition (SMI <1) while only 13.4% were in a better condition (SMI >1). There was no evidence of an association between body condition (Scaled Mass Index) and prevalence of *T. pestanai* in the badgers ($p = 0.739$, FET). The distribution of badgers has been studied in detail at Woodchester Park, for example, (Rogers et al. 1998). Badgers used in this study were raised in 28 setts (Table 4.4) and *T. pestanai* was detected in badgers

from 15 of these setts (Table 4.5). The majority of badgers were raised in Honeywell (13.4%) and Chestnut (7.3%) setts. Due to movements by some badgers, the badgers screened for this study now live in 26 setts and their distribution is shown in Table 3.5. All the badgers raised in Honeywell sett continued to live there and the sett dominates with the highest percentage (13.4%) of where the badgers now live. There was no statistical significant evidence suggesting association between where badgers were raised and their trypanosome infection status (p = 0.725, FET) neither was there statistical significant evidence indicating association between where badgers now live and their trypanosome infection status (p = 0.464, FET). As the sample size for each set was small and unevenness may be missed, setts were pooled into North and South regions to investigate any broad geographical unevenness. Fifty-two (63.4%) of the badgers live in the north, of these, 18 were infected resulting in 34.6% prevalence while the remaining 30 (36.6%) live in the south and 11 were infected resulting in 36.7% prevalence. The odds of getting infected with the disease was slightly higher for badgers living in the south than in the north (OR = 1.094; 95%; CI: 0.428 – 2.791) although this was not significant (P = 1.0). There was no evidence to suggest unevenness in the distribution of infection with respect to the parameters available.

Table 4.4. Showing setts where badgers were raised and trypanosome infection

Where badger was raised	Badger Population frequency		Trypanosome infection status		
	Frequency	Percent	not infected	infected	Total
beech	1	1.2	1	0	1
breakheart	1	1.2	1	0	1
box	2	2.4	2	0	2
boxwood	1	1.2	1	0	1
cedar	5	6.1	2	3	5
chestnut	6	7.3	3	3	6
colliers wood	1	1.2	1	0	1
evergreen	1	1.2	1	0	1
honeywell	11	13.4	8	3	11
inchbrook	5	6.1	4	1	5
jacks	4	4.9	2	2	4
junction	4	4.9	2	2	4
kennel	3	3.7	1	2	3
larch	4	4.9	3	1	4
mead	3	3.7	2	1	3
nettle	1	1.2	0	1	1
old oak	4	4.9	2	2	4
parkmill	5	6.1	2	3	5
poplar	1	1.2	1	0	1
scotland bank	2	2.4	0	2	2
west	5	6.1	5	0	5
windsoredge	5	6.1	3	2	5
woodfarm	1	1.2	1	0	1
woodrush	1	1.2	0	1	1
wychelm	1	1.2	1	0	1
yew	2	2.4	2	0	2
bracken	1	1.2	1	0	1
top	1	1.2	1	0	1
Total	82	100	53	29	82

Table 4.5. Showing setts where badgers now live and trypanosome infection

Where badger now lives	Badger Population frequency		Trypanosome infection status		Total
	Frequency	Percent	not infected	infected	
arthurs	2	2.4	2	0	2
beech	7	8.5	4	3	7
breakheart	1	1.2	1	0	1
boxwood	1	1.2	1	0	1
cedar	5	6.1	2	3	5
colepark	1	1.2	1	0	1
colliers wood	1	1.2	1	0	1
honeywell	11	13.4	8	3	11
inchbrook	5	6.1	4	1	5
jacks	4	4.9	2	2	4
junction	4	4.9	2	2	4
kennel	3	3.7	1	2	3
larch	6	7.3	5	1	6
mead	1	1.2	0	1	1
old oak	4	4.9	2	2	4
parkmill	5	6.1	2	3	5
scotland bank	2	2.4	0	2	2
West	5	6.1	5	0	5
windsoredge	5	6.1	3	2	5
woodfarm	1	1.2	1	0	1
woodrush	1	1.2	0	1	1
wychelm	2	2.4	2	0	2
yew	2	2.4	2	0	2
septic	1	1.2	1	0	1
top	1	1.2	1	0	1
nettle	1	1.2	0	1	1
Total	82	100	53	29	82

4.4 Discussion

This study records a 35.4% (25.9% - 46.2%; 95% CI) prevalence of *T. pestanai* infection in British badgers from Woodchester Park, Gloucestershire, UK. *T. pestanai* has been previously detected in British badgers from Cornwall (Pearce and Neal 1973) and Wytham Woods, Oxfordshire (Lizundia et al. 2011, Macdonald et al. 1999), with prevalences recorded at 10%, 7.7% and 31% respectively. Estimating an absolute prevalence rate may depend on the sensitivity of the diagnostic measure applied to detect trypanosomes in a sample population. For example, the studies that were based on detecting trypanosomes in blood smears from British badgers (Peirce and Neal, 1973 and Macdonald et al. 1999) recorded the lower prevalence (10% and 7.7% respectively) and this may be attributed to the difficulty in detecting trypanosomes in blood smears especially when the level of parasitaemia is very low. While a direct comparison cannot be made, the two studies using PCR-based methods (Lizundia et al. 2011) and this study showed a higher prevalence than those using microscopical analysis (Macdonald et al. 1999, Pearce and Neal 1973) suggesting a higher degree of sensitivity using the molecular approaches. The other PCR-based study (Lizundia et al. 2011), from a different site showed a prevalence of 31% (25.6 to 37.1; 95% CI) which was not significantly different to this study 35.4% (25.9% - 46.2%; 95% CI). The ITS-Nested PCR applied in our study appears to have comparable sensitivity to the 18S PCR (Lizundia et al. 2011) but has the added advantage of the potential to be used in any mammalian wildlife species and potentially detect any trypanosome species. Thus, offering a broader based tool for trypanosome surveys in wildlife.

However, prediction of expected band sizes for trypanosomes using the ITS Nested PCR is dependent upon the availability of the trypanosome rRNA gene regions (18S, ITS region, 5.8S and 28S) on the database. So it could prove a difficult challenge to predict a band size for a new species of trypanosome or where the rRNA gene regions (18S, ITS region, 5.8S and 28S) for the trypanosome is absent in the databases (e.g. *T. pestanai*).

Additionally, despite the potential of the ITS Nested PCR (Cox et al. 2005) for generic use in wildlife species, sequencing of the 18S rRNA gene, available on databases, needs to be used initially to confirm the trypanosome species or identify possible new species. One limitation of the 18S PCR, however, is due to its inability to distinguish when a sample is positive for mixed species infection with trypanosomes. In such cases it might become necessary to clone the 18S product and then screen the colonies with PCR-RFLP to distinguish between amplicons.

Additionally, the 18S (SSU) rRNA, and to a lesser degree the 28S (LSU) rRNA, sequence can be used to investigate phylogenetic relationships. On the SSU-rRNA tree, the badger trypanosomes detected in this study clustered with *T. pestanai* (100% bootstrap value), confirming their identity, and trypanosomes belonging to the *Megatrypanum* clade (80% bootstrap value). However, the addition of 28S rRNA sequence of the badger trypanosome did not improve the strength of the *Megatrypanum* clade and neither did we observe an improvement on any of the other major clades (Fig. 4.19). Concatenation of multiple sequences from the same taxa has been shown to improve phylogenetic inference in some studies (Gadagkar et al. 2005, Murphy et al. 2001, Soltis et al. 1999).

However, there are recent reports of how phylogenetic reconstruction using concatenated datasets have led to poor phylogenetic tree estimates (Kubatko and Degnan 2007, Salichos and Rokas 2013). In our analysis, the concatenated tree had some missing datasets due to the

unavailability of some trypanosome 28S rRNA in the sequence databases. The possibility exists that this may have altered the topology of our tree. However recent studies, using simulations, suggest there is little evidence to support poor resolution of phylogenetic trees from the exclusion of taxa due to missing data when using concatenated sequences as long as many characters have been sampled overall (Wiens 2006). In this study, the constructed phylogenetic trees appeared to be robust. The *Schizotrypanum* clade (Da Silva et al. 2004) was strongly supported with 89% bootstrap support on SSU-rRNA tree (Fig. 4.18) and 87% support on the concatenated tree (Fig. 4.19). Additionally, both trees confirm *T. rangeli* to be more closely related to the *Schizotrypanum* clade than to the *Herpetosoma* clade which is also consistent with the phylogenetic analysis from previous studies (Da Silva et al. 2004, Stevens et al. 1999). Trypanosomes belonging to the *Herpetosoma* subgenus were also shown to cluster together on both trees with a high bootstrap support of 100% in SSU-rRNA tree (Fig. 3.17) and 99% in concatenated tree (Fig. 4.19). Also, fish trypanosomes (*T. carpio and T. ophiocephali*) were shown to be monophyletic in previous research (Gu et al. 2010). Phylogenetic analysis from this study deduced using ML also shows both species to form a sister clade with a high bootstrap support value in both trees (100 % SSU-rRNA, Fig. 4.18; 100% concatenated tree, Fig. 4.19). The amphibian trypanosomes (*T. mega and T. rotatorium*) (Bardsley and Harmsen 1973) included in this analysis, however, did not cluster together on both trees (Fig. 4.18 and Fig. 4.19). Salivarian trypanosomes were shown to cluster together in both trees with a strong bootstrap support of 100%. Also, both trees show *T. evansi* to be closely related to the human infective forms of *T. brucei* with a high bootstrap support value of 100% (Fig. 4.18 and Fig. 4.19).

In conclusion, the use of nested ITS-PCR for the detection of *T. pestanai* in British badgers shows that it has potential to be an effective tool for molecular detection and identification of trypanosomes in wildlife. Although mixed species infections were not detected in these

samples, in the generic case of trypanosome detection in wild animals, unlike the 18S PCR, this approach offers the potential to identify such mixtures if they existed. Such mixtures have been identified in African cattle (Cox et al. 2010). The observed prevalence of 35.4% in this study and comparable figures from related studies (Lizundia et al. 2011) suggests that a significant proportion of UK badgers could be infected with trypanosomes. As with many host-parasite systems, this shows that trypanosome infection may be a normal arrangement in these hosts. The Hygiene Hypothesis e.g. (Daley 2014) has been often quoted in humans as being responsible for our increased susceptibility to allergies and ailments. Perhaps the natural infection with parasites, such as trypanosomes, may influence the badger immune system and help in the resistance to other important pathogenic diseases, such as tuberculosis. More studies are required to investigate the general transmission of infectious diseases in this, and other, important potential wildlife reservoirs.

Chapter 5: Genetic and epigenetic variation of the bovine TLR9 gene

5.1. Introduction

Mammals possess toll-like receptors (TLRs) which are major defenders of our innate immune system. Ten bovine TLRs have been identified to be homologous to the ten TLRs found in humans. They play a crucial role in bovine innate immune system to recognise molecular signatures of invaders known as pathogen-associated molecular patterns (PAMPs) which are usually not expressed by the host (Medzhitov and Janeway 2000, Menzies and Ingham 2006). However, genetic alterations in TLR genes are increasingly implicated to be associated with infectious diseases. For example, Mucha et al. (2009) identified that two missense mutations in TLR4 are associated with *Mycobacterium paratuberculosis* MAP infection in cattle. Also, TLR4 polymorphisms have been shown to be associated with malaria manifestation in African children (Mockenhaupt et al. 2006). There are increasing reports of single nucleotide polymorphisms (SNPs) found in TLR genes associated with protozoa infections (Wujcicka et al. 2013).

In protozoa infections, the glycosylphosphatidylinositol (GPI) - anchored surface molecules are a major set of PAMPs that trigger inflammatory responses through TLR2 and TLR4 via the MyD88 signalling molecule (Gazzinelli and Denkers 2006). For example, TLR2 has been shown to recognise GPI anchors from *Trypanosoma cruzi* in humans (Campos et al. 2001) but in African trypanosomes the VSGs have been referred to as the major GPI-anchored surface structures (Mishra et al. 2009). Drennan et al. (2005) showed that *T. brucei*-derived soluble VSG (sVSG) and membrane-bound VSG (mVSG) activate macrophages in a MyD88 dependent manner and that macrophages deficient in TLR9 responded poorly to *T. brucei* genomic DNA, implying that trypanosome parasitaemia control is partly mediated by TLR9.

Interestingly, Harris et al. (2007) reported how enhancement of the immune response by artificial stimulation with synthetic cytosine-phosphate-guanine oligodeoxynucleotides (CpG ODN), a TLR9 agonist resulted in enhanced protection against African trypanosomiasis. Their work showed that in susceptible BALB/c mice, relatively resistance to trypanosome infection was significantly enhanced by CpG ODN treatment and associated decreased parasitaemia. While, in a relatively resistant C57BL/6 mice, survival was not enhanced, but parasitaemia levels were reduced. Thus, suggesting that resistance to African trypanosomes can be induced in susceptible animals in a TLR9- dependent manner.

Despite the accumulating evidence suggesting the association of the TLR9 gene with trypanosomes and also TLRs polymorphisms with infectious diseases, reports on the association of TLR9 polymorphisms and trypanosome infection is still lacking. In a recent study, on the genome-wide analysis of DNA methylation in bovine placentas, the bovine TLR9 gene was among one of the bovine genes found to be methylated in their analysis (Su et al. 2014). One method of identifying if a gene could be potentially methylated is by first identifying cytosine-phosphate-guanine Islands (CpG Islands) in the gene. The CpG Islands are sites in DNA sequence where there are clusters of CpG or GC which are known to be potential sites for DNA methylation (Taqi et al. 2011). This is then followed by bisulfite treatment of the DNA, which converts unmethylated cytosines to uracils. Methylated cytosines remains unchanged so after treatment methylation status of the gene could be confirmed using bisulfite PCR or methylation specific PCR (http://www.zymoresearch.com). Detailed procedure for the bisulfite conversion is described in Chapter 2.9 of this report. The identification of SNPs in CG sites of a methylated gene could play important epigenetic markers for studying the expression of the gene. In some cases, such SNPs have been associated with diseases. For example, epigenetic variants based on SNPs found in CpG

Islands have been shown be associated with type 2 diabetes (Dayeh et al. 2013) and risky behaviours such as alcohol dependence (Taqi et al. 2011).

Hence, the purpose of this study was targeted at investigating genetic and epigenetic variants in African bovine TLR9 genes, and to identify potential genetic and epigenetic markers which perhaps could be used to study infectious diseases. The study was also aimed to test the association of any identified marker with bovine samples of known trypanosome infection status. To achieve this, bovine samples of known trypanosome infection status were collected from collaborators at the University of Edinburgh, regions in bovine TLR9 gene covering the two CpG Islands were selected as targets for analysis in this study. The following objectives were then carried out;

- For genetic analysis, primers would be designed and applied to amplify the complete bovine TLR9 mRNA coding sequence region.
- Development and application of a novel hemi-nested PCR to amplify the target sequence covering the two CpG Islands before PCR products would be sent to GATC biotech a commercial sequencing company for purification and sequencing.
- Analysis of bovine TLR9 DNA sequences and comparison with published national centre for biotechnology information reference sequence for bovine (*Bos taurus*) TLR9 gene (NM_183081.1).
- For epigenetic analysis, first the DNA would be converted to bisulfite DNA using a commercial kit, and then bisulfite primers targeting the CpG Islands would be designed for bisulfite PCR. A hemi-nested bisulfite PCR would be developed and applied to amplify CpG Islands. PCR products would be sent to GATC biotech for purification and sequencing. Bisulfite sequence analysis would be carried to identify epigenetic variants.

5.2. Materials and methods

A total of 72 cattle blood samples from Nigeria and Uganda, collected using Whatman FTA cards, were available for the study. The samples were part of a long-term study on African trypanosomiasis conducted at the University of Edinburgh. In collaboration with Prof. Geoff Hide (My supervisor) and Dr. Kim Picozzi (University of Edinburgh), I was granted access to the research team in Edinburgh who gave me the samples which I collected as single FTA card punches for 57 Nigerian samples (57 punches, 1 per sample) and whole blood spotted FTA cards for the Ugandan samples (16 samples). The Nigerian samples were from three villages; Tambes (19), Bokkos (18) and Kachia (16) in northern Nigeria (Table 5.1).

Table 5.1. Cattle samples used for the study

	Nigeria			**Uganda**	**Total**
	Tambes	Bokkos	Kachia	Uganda	
	T1	B1	K1	U5	
	T2	B2	K2	U6	
	T3	B3	K3	U7	
		B4	K4	U8	
	T5	B5	K5	U9	
	T6	B6	K6		
	T7	B7	K7	U11	
	T8	B8	K8	U12	
	T9	B9	K9	U61	
	T10	B10	K10	U62	
	T11	B11	K11	U63	
	T12	B12	K12	U64	
	T13		K13	U85	
	T14	B14	K14	U86	
	T15	B15	K15	U87	
	T16	B16	K16	U88	
	T17	B17	K17	U93	
	T18		K18		
	T19	B19	K19		
	T20	B20			
Total	**19**	**18**	**19**	**16**	**72**

The samples were prepared using the FTA card preparation procedure described in chapter 2.2.2 of this report before the bovine DNA was eluted from the prepared single punch using

Chelex ® elution protocol (Chapter 2.2.3). Usually the FTA card allows for a direct amplification of prepared punches without further elution of DNA using chelex but our aim of DNA elution from the single punches using chelex was to obtain sufficient DNA for several rounds of PCR amplifications. This was because we could not guarantee success in every experiment and the situation may arise when we need to fall back to the same DNA for a repeated PCR re-amplification.

Firstly, a novel PCR for the amplification of the coding region of bovine TLR9 gene using genomic bovine DNA was developed. The newly developed bovine TLR9 coding region PCR was then tested on FTA card samples. The EMBOSS Cpgplot (https://www.ebi.ac.uk/Tools/seqstats/emboss_cpgplot/) was used to identify the TLR9 CpG Islands which became a target for a Hemi-nested PCR to increase the quality of DNA and subsequently derive sequence data for the region.

The PCR products from all bovine samples previously amplified using the bovine TLR9 coding region PCR were then used in the Hemi-nested approach to derive DNA sequence data for areas covering the bovine TLR9 CpG islands. The PCR products derived were sourced out to GATC Biotech Ltd. (www.gatc-biotech.com), a commercial sequencing company for purification and sequencing.

Furthermore, to carry out epigenetic analysis of the bovine TLR9 DNA, 20 bovine samples eluted using the chelex elution procedure (chapter 2) were converted to bisulfite DNA using EZ methylation gold kit from Zymo Research (www.zymoresearch.com) according to manufacturer's recommendation. Methprimer a programme for designing bisulfite-conversion-based methylation primers (www.urogene.org/methprimer) was used to design primers for bisulfite PCR. Novel bisulfite hemi-nested PCR for amplification of bovine TLR9 CpG Islands was developed and applied to bisulfite treated bovine DNA. PCR products were sent out to GATC biotech and sequence data was analysed visually and with the aid of Finch

TV for comparison with previous genetic data derived from sequencing. All statistical analysis for this study were carried out using IBM SPSS Statistics for Windows (version 20).

5.3. Results

5.3.1. Development of TLR9 coding region PCR

Firstly, DNA sequences for the bovine TLR9 gene were searched on the NCBI nucleotide database which archives a collection of sequences from several sources, including GenBank, RefSeq, third-party annotation (TPA) and protein data bank (PDB). The NCBI reference DNA sequence for *Bos taurus* TLR9 gene was identified and was shown to be well annotated but was 3265bp (NM_183081.1) in size while another *B. taurus* sequence had a genomic size of 5033bp (EF076731) so this latter was chosen for primer design purposes. A multiple sequence alignment of bovine TLR9 genes was made using the Cluster Omega software (www.ebi.ac.uk/Tools/msa/clustalo/). This showed that the *B. taurus* species were similar in sequence (see Appendix G) and so the DNA sequence with accession number (EF076731) which is bigger in size was selected for primer design because it contained 5033bp and was well annotated to include the coding region which was a target for PCR amplification. Using this sequence annotated on NCBI, the bovine TLR9 mRNA was shown to have to two exons joined by an intron and forming a coding sequence region (Fig. 5.1).

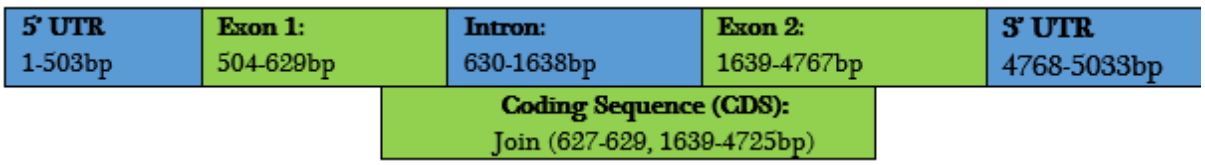


Figure 5.1. Illustration of bovine TLR9 gene (Accession: EF076731). 5' UTR (1-503bp) is shown in blue colour, mRNA (Exon1; 504 to 629bp and Exon2; 1639 to 4767bp) coding sequence region (627 to 629bp joining 1639 to 4725bp) is coloured green. While the intron (630 to 1638bp) and 3' UTR (4768 to 5033bp) is coloured blue.

Therefore, the TLR9 primers were designed to cover the coding region. The primers were selected after alignment with cluster omega and the Primer-BLAST programme (www.ncbi.nlm.nih.gov/tools/primer-blast/) which is a tool for designing primers was used to test for primer sensitivity and specificity which confirmed our expected band size to be 4156bp. The TLR9 forward primers sequence is; 5' – GGA GAA GCC GCA TTC CCT G - 3' while the reverse primer sequence is; 5' – TGT GGG GTT AAA GGA GTG CTG- 3' with the 5' for both primers located at position 606 and 4761bp respectively (Accession no: EE076731). The PCR was carried out using the protocol described in Chapter 2.7 of this report. To test the newly designed TLR9 primers, the first step was to acquire bovine DNA to be used as the control sample. So beef was purchased at a local store (Sainsbury) and DNA was extracted using the phenol-chloroform procedure described in Chapter 2.2.1 of this report. The DNA was then used for TLR9 coding region PCR. Unfortunately, the primers could not amplify regions of bovine TLR9 gene after several attempts. All reagents used for the PCR amplification (as specified in chapter 2.7) were changed for new ones except for the primers. Again there were no visible bands. Since, there was not even a single band our suspicion was that the extraction had failed. However, following repeated extraction and PCR amplification using the same procedure as before there was still no visible bands (Fig. 5.2).

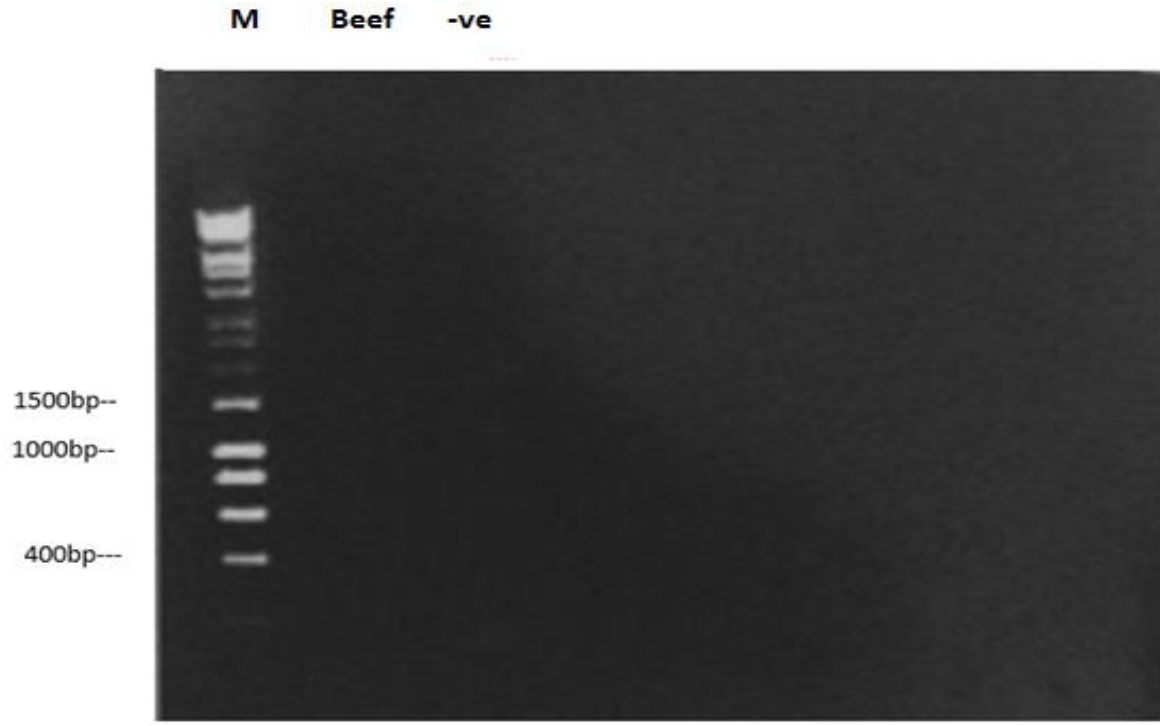


Figure 5.2. Failed bovine TLR9 coding region PCR. No visible bands on beef lane indicate failed amplification. M is hyperladder 1 marker while –ve is water used as a control.

To save time and cost, bovine genomic DNA was purchased from a commercial company Amsbio (http://www.amsbio.com) which was used to test the bovine TLR9 coding region PCR using reaction conditions detailed in Chapter 2. Two PCR reaction ready-to-use mixtures (RANGER mix and Myfi ™ mix) were purchased from Bioline based on their capability to amplify genomic DNA up to 25kb and 10kb respectively. Initial amplification of the newly purchased DNA using both ready-to-use mixtures according to the manufacturer's recommendation showed nonspecific bands (Fig. 5.3).

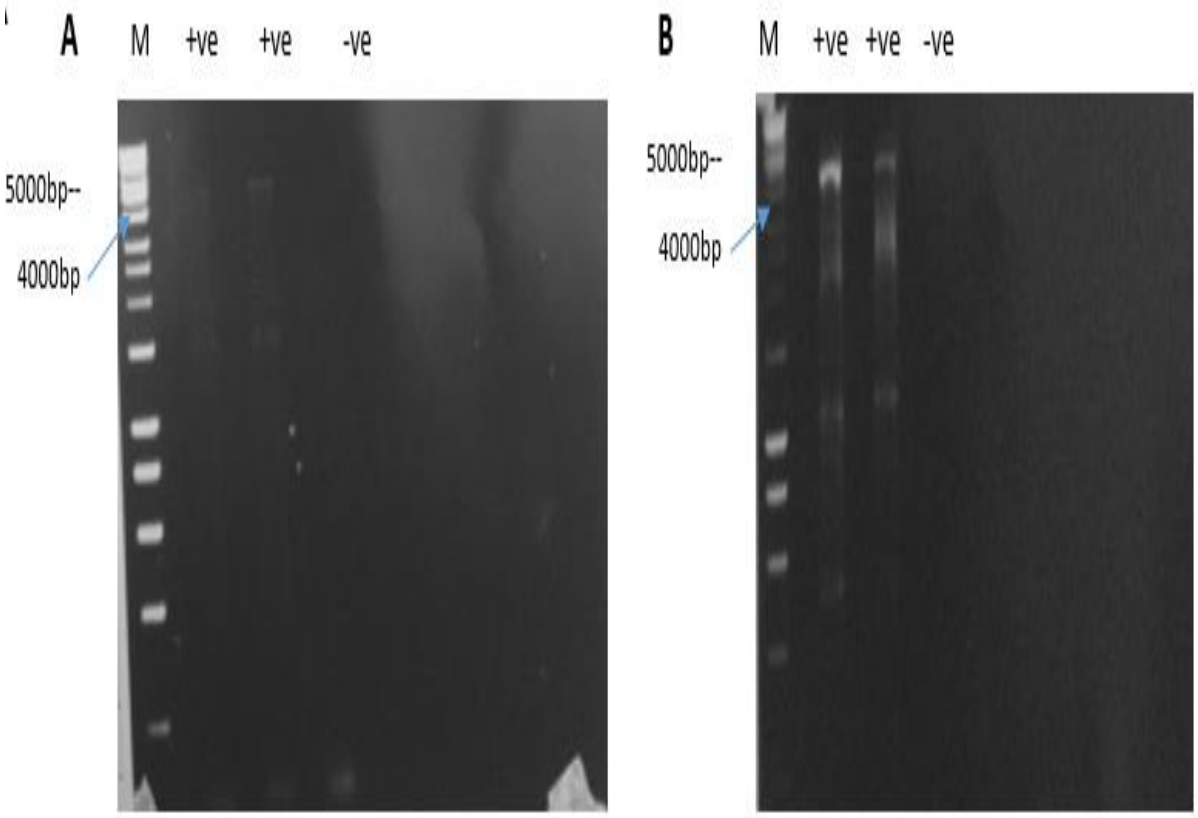


Figure 5.3 Development of TLR9 coding region PCR using RANGER mix and Myfi ™ mix. Gel image A shows amplification using Myfi ™ mix while gel image B shows amplification using RANGER mix. In both images, M represents hyperladder 1kb plus, +ve are bovine DNA from Amsbio while negative is water used as a control sample.

However, this was promising as it indicates that the DNA can be amplified by PCR using the newly designed TLR9 primers suggesting optimisation might result in more specific bands. So, a gradient PCR at 55° C, 58° C and 60° C using RANGER and Myfi ™ mix was carried out to determine the optimal annealing temperature. The results of the PCR amplification showed clear bands of 4156bp at 55° C from DNA amplified in both reaction mixtures (Fig. 5.4), marking a successful development of bovine TLR9 coding region PCR.

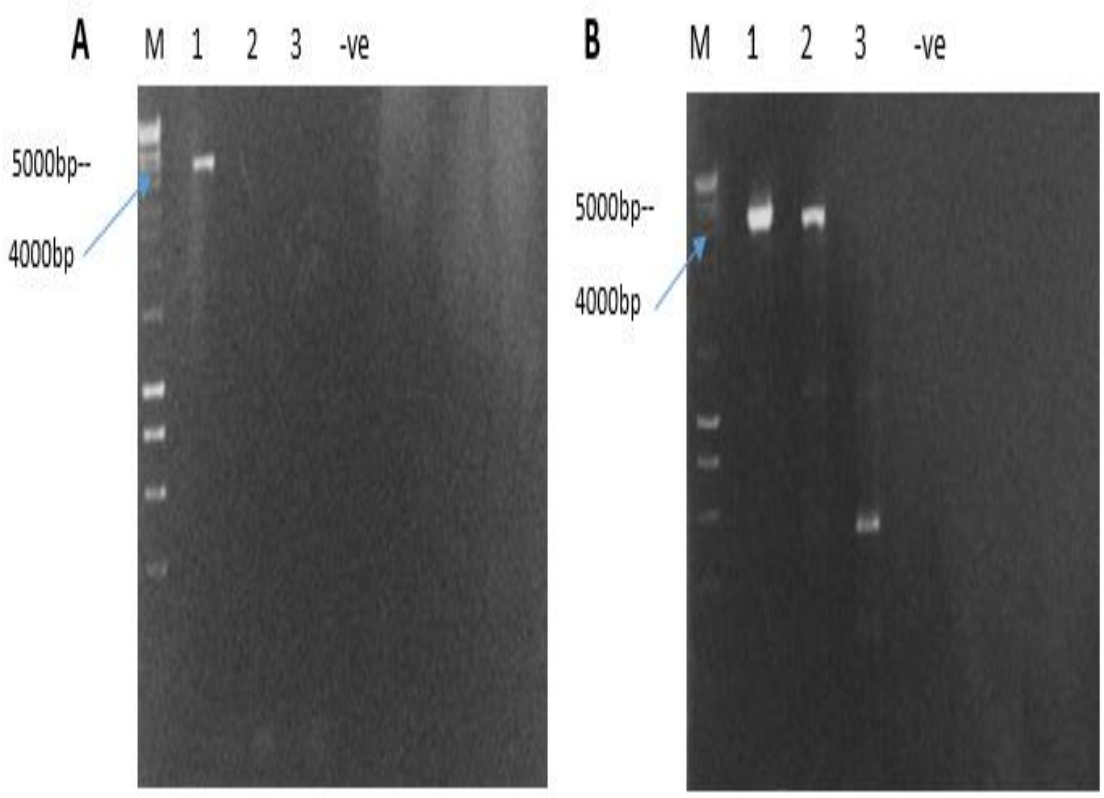


Figure 5.4 TLR9 coding PCR using RANGER mix and Myfi ™ mix. Gel image A shows amplification using Myfi ™ mix while gel image B shows amplification using RANGER mix. In both images M represents hyperladder 1kb plus, lane 1, 2 and 3 are bovine DNA from Amsbio set at gradient temperatures (55^0 C, 58^0 C and 60^0 C respectively) while negative is water used as a control sample.

Following the successful development of TLR9 coding region PCR, the technique was tested in some bovine samples (AH32, AH33, AH34, AH35 and AH36) collected on Whatman FTA used in previous studies. This aim was to send the PCR products for sequencing and ensure that the technique can yield good sequencing data before applying it to the original samples for this study collected from Edinburgh University. Amplification of the test samples produced faint bands, implying that the TLR9 coding region PCR also works on samples from FTA cards (Fig. 5.5).

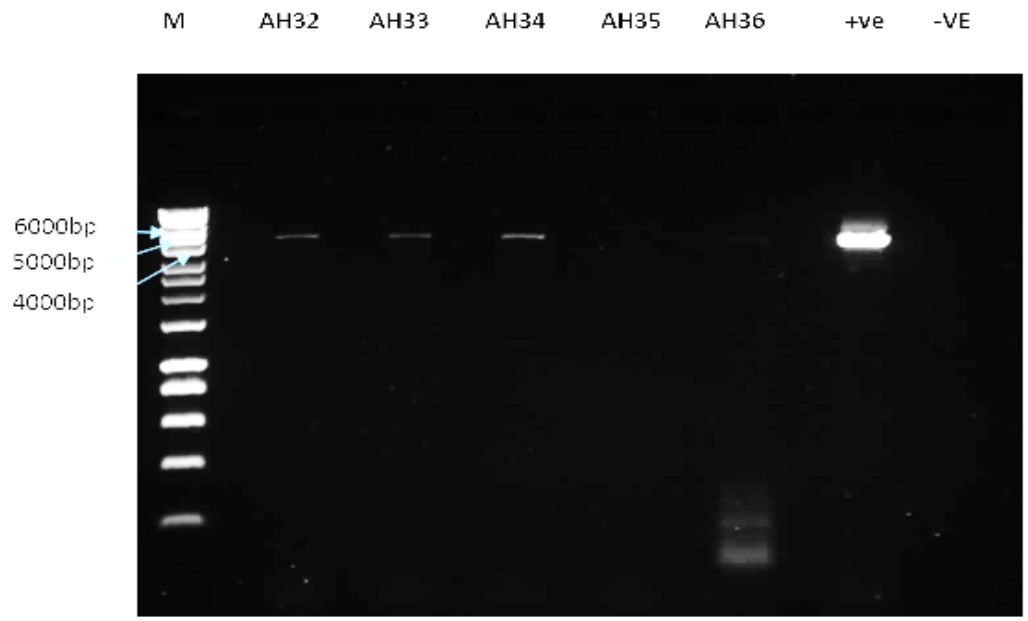


Figure 5.5. TLR9 coding region PCR using FTA cards. Lane M represents marker of known molecular weight. Lane AH32-AH36 are FTA cards samples. All bovine FTA card samples produced faint bands but AH35 did not produce a band. +ve represents a known bovine sample and –ve is water.

PCR products for 2 of these samples together with the control sample were sent to GATC biotech Ltd. for purification and sequencing according to the commercial company's sequencing requirements. The sequence data received from the company was opened using a DNA sequencing analysing software (Finch TV). Unfortunately, the sequencing of PCR products from the FTA cards did not yield any sequence data while the control sample (Amsbio bovine DNA) yielded a good sequence data but analysis using the Megablast software programme on NCBI that searches for homologous sequences showed that the sequence derived from TLR9 coding region PCR did not cover all areas of the bovine TLR9 coding region. Enquiries were made to the technical team at GATC who advised that their reads cannot go above 1200bp on normal Sanger sequencing and that they could not provide sequence data for the FTA cards products because the DNA concentration was low. Their advice was to design walking primers to cover all areas of the TLR9 genes if interested. However, on consideration that the bands from FTA cards were faint, we targeted a partial

amplification of the bovine TLR9 gene with a main focus on the amplification of the CpG Islands. Detailed procedures for TLR9 coding region PCR is available in Chapter two.

5.3.2. Identification CpG Islands and Development of bovine TLR9 Hemi-nested PCR

Using the EMBOSS Cpgplot (https://www.ebi.ac.uk/Tools/seqstats/emboss_cpgplot/), we detected 2 CpG islands from bovine TLR9 genomic sequence (Accession no: EF076731). Bovine TLR9 was reported to have these islands while TLR3, 7 and 8 were reported to have no CpG Islands (Cargill and Womack 2007), also no CpG Islands were observed in bovine TLR1, 5 and 10 (Seabury et al. 2007). So, using CpgPlot from EMBOSS we investigated bovine TLR2 (Accession: NM_174197.2), TLR4 (Accession: NM_174198.6) and TLR6 (Accession: NM_001001159.1) but also detected no CpG Islands in these TLRs (See appendix F for CpG graph plot). Thus, suggesting that amongst all 10 bovine TLRs CpG Islands was detected in TLR9 within the coding sequence region using the prediction EMBOSS Cpgplot tool.

Nonetheless, the first CpG island detected in bovine TLR9 has a length of 259bp and is located from position 3721-3979bp while the second CpG island has a length of 421 and is located from 4182-4602bp (Fig. 5.8). Two primers were selected and designed to cover these regions. The forward primer; TLR9CpGF (5'–CTGCGTCTCCGGGACAATAAC-3') starts at 3601bp while the reverse primer TLR9CpGR (5'CGGTTATAGAAGTGACGGTTG-3') starts at 4691bp. The primers were used as internal primers to amplify PCR products derived from TLR9 coding region PCR. The MyTaq™ Hs ready-to-use reaction mixture purchased from Bioline was used (according to the manufacturer's recommendation) for a nested PCR to amplify PCR products derived from TLR9 coding region PCR. Unfortunately, the results

after gel electrophoresis showed nonspecific bands (600bp) indicating the technique had not worked (Fig. 5.6).

The experiment was repeated several times trying different optimisation procedures such as increasing and/or reducing cycles and extension times and changing reagents (RANGER mix, Myfi ™ mix and MyTaq ™ HS mix) but the unspecific bands of 600bp continued to appear and remained the same as shown in Fig 5.6. Various temperature gradients were carried, but this could not solve the problem. In some cases, the band would appear very faintly (eg T11, Fig. 5.7) but when the experiment was repeated at the same temperature even with the same sample the nonspecific band reappears.

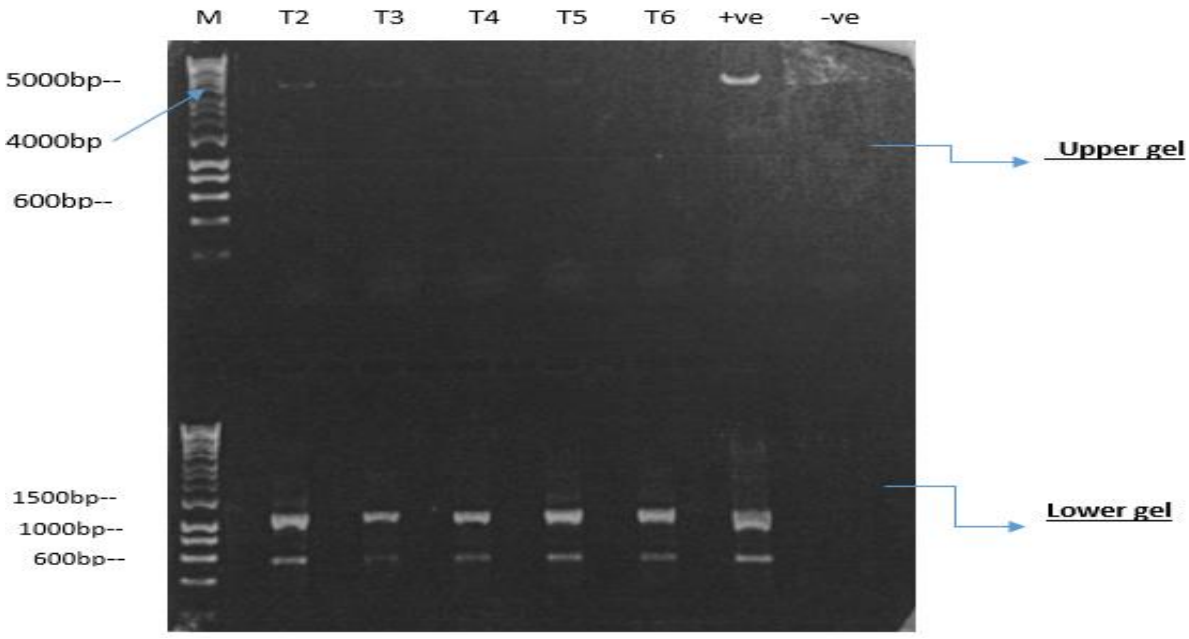


Figure 5.6. Failed nested PCR using TLR9CpGF and TLR9CpGR. The upper gel represents 1st round of PCR (TLR9 coding region PCR) while lower gel is 2nd round of PCR (TLR9 CpG PCR). PCR products from 1st round (upper gel) were used for amplification in 2nd round (lower gel). T2 to T6 are study samples. Double bands on samples indicate failed PCR. +ve is Amsbio bovine control DNA while –ve is water.

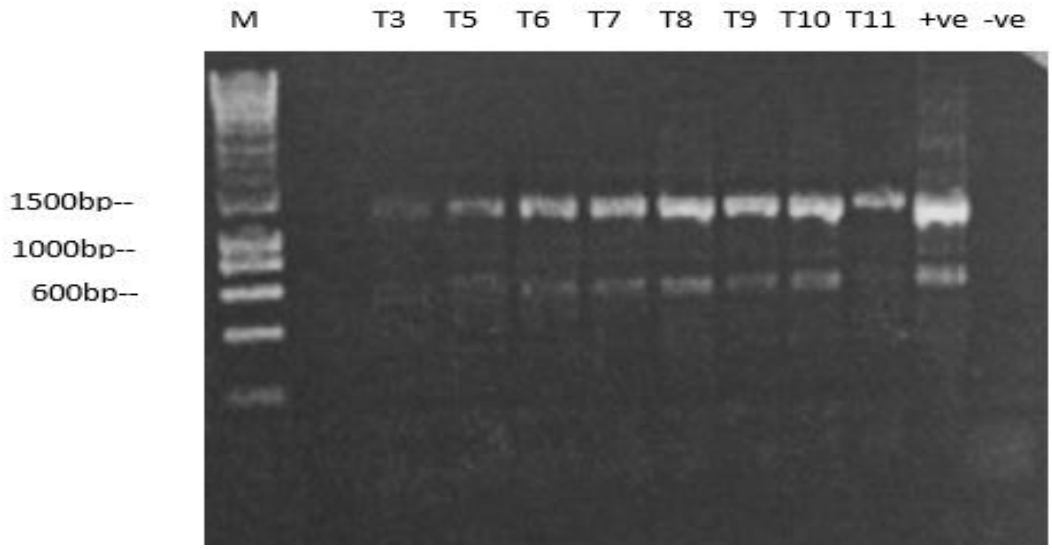


Figure 5.7. Failed gradient TLR9 CpG PCR. T3-T11 are study samples used at gradient temperatures (from 52^0 C to 59^0 C). Double bands on samples indicate failed PCR. M is hyperladder 1kb plus marker while +ve is Amsbio bovine control DNA amplified at 60^0 C and –ve is water.

Having exhausted most optimisation steps for PCR amplification, the next step was to apply a Hemi-nested PCR approach. The hemi-nested PCR involves the amplification of a target sequence using two consecutive PCR rounds, in which the first round uses the first set of external primers while the second round uses a new internal primer and either of the two primers from the first round of PCR. It has been applied in various studies (Bachman et al. 2004, Djordjevic et al. 1998, Pujol-Rique et al. 1999) based on its high sensitivity and specificity to amplify target sequences. For this study, the hemi-nested PCR used the two primers in TLR9 coding region PCR (TLR9F and TLR9R) for first round amplification and the second round was carried using forward primer (TLR9 CpGF) from TLR9 CpG PCR and the reverse primer (TLR9R) from the first round (Fig. 5.8).

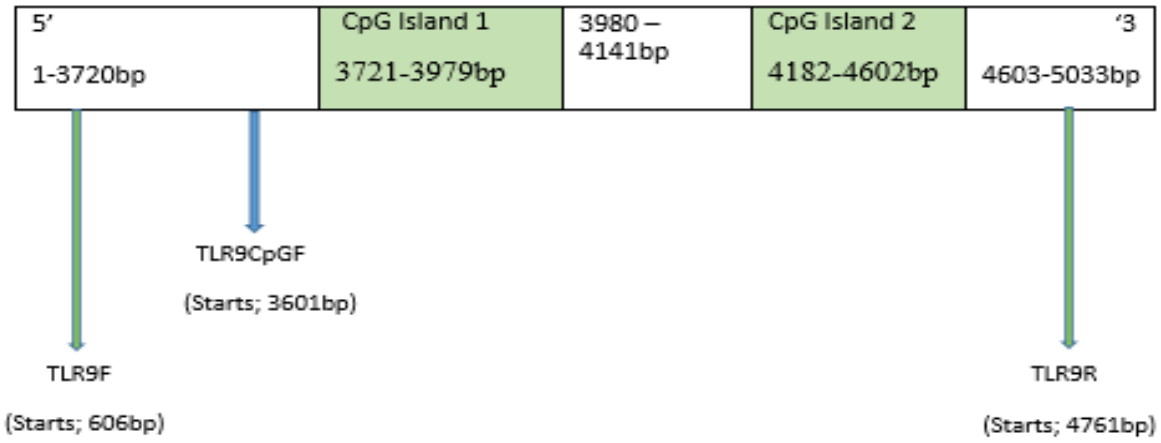


Figure 5.8. Illustration of bovine TLR9 hemi-nested PCR. The representative *Bos taurus* TLR9 mRNA (Accession no: EF076731). Areas shaded green are the CpG islands which are the target for amplification. Green arrows are primers (TLR9F and TLR9R) used in first round PCR. The blue arrow is the internal forward primer (TLR9CpGF) used with the TLR9R primer for the second round of PCR.

The expected band size for the TLR9 hemi-nested PCR is 1360bp. The reaction conditions are detailed in Chapter two of this report. The technique was first tested with 4 samples (T1, T3, U5 and Amsbio bovine DNA). The result after gel electrophoresis showed the estimated band sizes of 1360bp in all four lanes indicating the amplification of the target sequence was successful, although a faint band was produced in lane U5 it was rather due to pipetting errors when loading the sample in the agarose gel (Fig. 5.9).

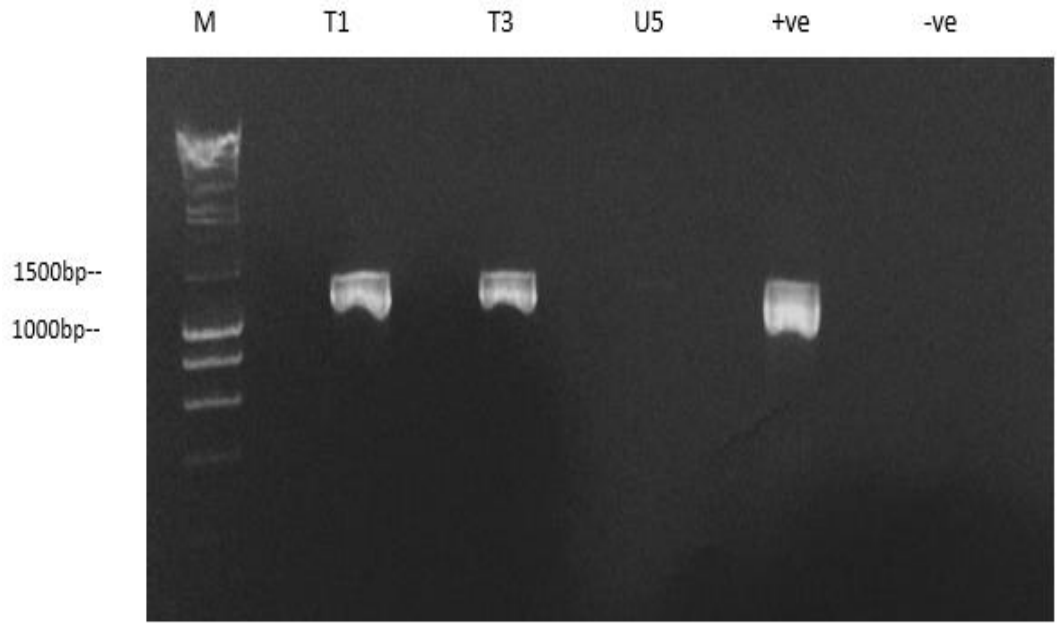


Figure 5.9. Gel image of successful hemi-nested PCR. Appearance of the estimated band sizes of 1360bp in samples T1, T2, T3 and +ve control (Amsbio bovine DNA) indicate the development of the technique has been successful. M is hyperladder 1kb plus while –ve is water used as a control.

Following the development of the technique, all 72 DNA bovine samples used in the study were amplified using the hemi-nested PCR approach in view of deriving bovine TLR9 partial coding sequence data for areas covering the two predicted CpG islands. The hemi-nested PCR always resulted in good quality amplifications and there was never a problem with contamination. All PCR products after hemi-nested PCR for bovine samples were sent to GATC for purification and sequencing.

5.3.3 Bioinformatics and statistical analysis of TLR9 genetic variants

The PCR products from the 72 bovine samples which were sent to GATC for sequencing were received electronically and analysed using the computer software programme Finch TV, which showed very clear peaks (Fig. 5.10).

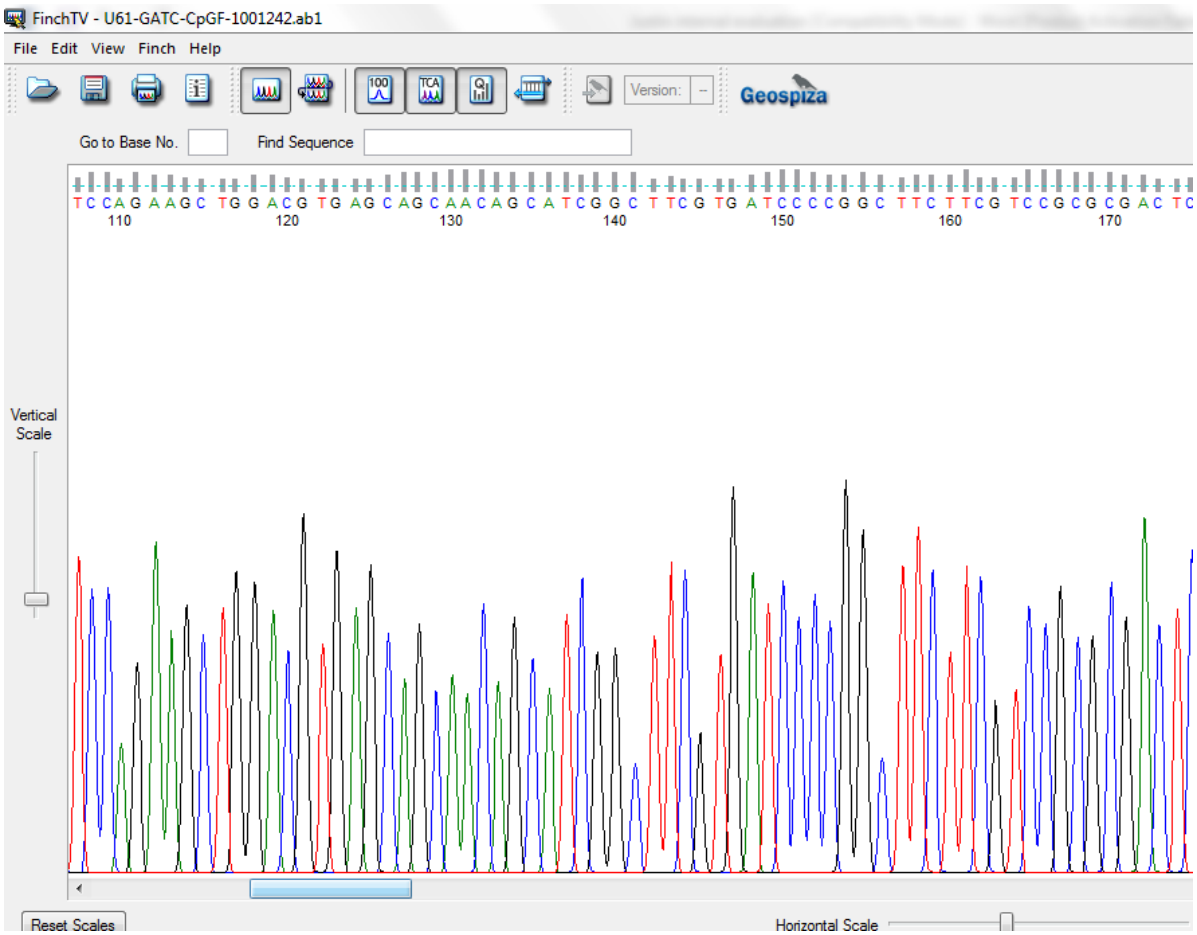


Figure 5.10. Finch TV analysis of bovine TLR9 hemi-nested PCR sequence data. The clear peaks shown are representative of nucleotide bases. Each nucleotide is denoted by a colour (Green represents adenine, black represents guanine, blue represents cytosine and red represent thymine.

The DNA sequence data, when retrieved, were used to search for highly similar sequences in NCBI database using the MegaBlast software programme. The results shows alignment with bovine TLR9 genes in all cases but there were variations in some specific loci which were continuously observed in the sample population, most of these were identified as single nucleotide polymorphisms with known references on the NCBI SNP database while some were a constant and appeared to be species-specific.

5.3.3.1 SNPs Identification

DNA sequence data for bovine TLR9 was derived from 72 bovine samples. The *B. taurus* TLR9 reference gene with DNA sequence (NM_183081.1) and protein sequence (NP_898904.1) available on NCBI database were used to compare all sequences derived from this study. In total, nine polymorphisms (SNP 2238, 2256, 2538, 2547, 2583, 2823, 2865, 2916 and 2924) were observed in the samples on analysis of the derived TLR9 sequence data (Table 5.2). Full details of polymorphisms observed per individual sample are available (Appendix H).

Table 5.2. Derived bovine TLR9 SNPs

SNP Nucleotide Position (bp) NM_183081.1	SNP Nucleotide positon (bp) EF076731	NCBI SNP Reference	Novelty	Observed in study samples?	Observed in Reference *B. indicus*?	Observed in reference *B. taurus*?	SNP Allele	Amino acid change	Amino acid position
2238	3750	N/A	Yes	Yes	Yes	No	C/G*	D/E	705
2256	3768	rs55617244	No	Yes	Yes	Yes	C/T	No	711
2538	4050	rs209982925	No	Yes	Yes	Yes	C/T	No	805
2547	4059	rs518414196	No	Yes	Yes	Yes	A/G	No	808
2583	4095	rs55617221	No	Yes	Yes	Yes	A/G	No	820
2823	4335	rs55617235	No	Yes	Yes	Yes	G/C	No	900
2865	4377	rs55617220	No	Yes	Yes	Yes	A/G	No	914
2916	4428	N/A	Yes	Yes	Yes	No	C/T*	No	931
2924	4436	N/A	Yes	Yes	Yes	No	A/C*	D/A	734

*Constant. For example at position nucleotide 2238, there is always a cytosine in derived DNA sequences but guanine in TLR9 reference gene (NM_183081.1)

*N/A denotes reference not available on NCBI database.

Also, analysis of derived sequence data showed three SNPs (SNP 2238, 2916 and 2924) to be constant in all samples so they were excluded from most parts of the statistical analysis. The

three SNPs which were constant in all our samples were also present in all other published *B. indicus* and *B. indicus* sequences on NCBI database but were absent in only two sequences (AJ509824.2 and AJ509825.1) which are similar to the *B. taurus* RefSeq (NM_183081.1). However, details from NCBI shows the RefSeq was derived from these sequences which were published by the same author Goldammer et al. (2004). Based on the dissimilarity of the RefSeq with other sequences and ours, it is difficult to ascertain if these could be regarded as actual SNPs or might have occured due to sequencing errors. The nucleotide differences observed between the RefSeq (NM_183081.1), a published sequence subspecies our sequence (TI) is shown in Fig. 5.11 below.

Figure 5.11 Alignment of derived TLR9 sequence with published sequences. NM_183081.1 is DNA sequence from published *B. taurus* TLR9 gene, T1 is derived sequence from this study while FJ495080.2 is a published *B. indicus* TLR9 sequence. Colours indicate nucleotide positions of polymorphism. Green shows polymorphism at position 2238, blue shows polymorphism at nucleotide position 2916 while red shows polymorphism at nucleotide position 2924.

As shown in Fig. 5.11 above, at nucleotide position 2238 *B. taurus* (NM_183081.1) possess a guanine nucleotide while *B. indicus* (FJ495080.2) and our derived TLR9 sequence (T1) possess a cytosine nucleotide. Also, at position 2916 *B. taurus* had thymine while T1 and *B. indicus* possess a cytosine. Lastly, at nucleotide position 2924 *B. taurus* possessed a cytosine whereas T1 and *B. indicus* possess an adenine. Only the observed polymorphism at position 2238 and 2924 are non-synonymous causing amino acid changes from aspartic acid (D) to

glutamic acid (E) and aspartic acid (D) to alanine (A) at amino acid position 705 and 934 respectively. The bovine TLR9 protein ID is NP_898904.1 and the amino acid changes observed are shown in the alignment in Fig. 5.12 below.

```
T1              ----------------------------ALSNGSLPPGIRLQKLDVSSNSIGFVIPGFFV      32
NM_183081.1     NNLAFFNWSSLTVLPRLEALDLAGNQLKALSNGSLPPGIRLQKLEVSSNSIGFVIPGFFV      720
                                           ****************:***************

T1              RATRLIELNLSANALKTVDPSWFGSLAGTLKILDVSANPLHCACGAAFVDFLLERQEAVP      92
NM_183081.1     RATRLIELNLSANALKTVDPSWFGSLAGTLKILDVSANPLHCACGAAFVDFLLERQEAVP      780
                ************************************************************

T1              GLSRRVTCGSPGQLQGRSIFTQDLRLCLDETLSLDCFGLSLLMVALGLAVPMLHHLCGWD      152
NM_183081.1     GLSRRVTCGSPGQLQGRSIFTQDLRLCLDETLSLDCFGLSLLMVALGLAVPMLHHLCGWD      840
                ************************************************************

T1              LWYCFHLCLAHLPRRRRQRGEDTLLYDAFVVFDKVQSAVADWVYNELRVQLEERRGRRAL      212
NM_183081.1     LWYCFHLCLAHLPRRRRQRGEDTLLYDAFVVFDKVQSAVADWVYNELRVQLEERRGRRAL      900
                ************************************************************

T1              RLCLEERDWLPGKTLFENLWASVYSSRKTMFVLDHTDRVSGLLRASFLLAQQRLLEDX--      270
NM_183081.1     RLCLEERDWLPGKTLFENLWASVYSSRKTMFVLAHTDRVSGLLRASFLLAQQRLLEDRKD      960
                ********************************* ***********************
```

Figure 5.12 Alignment of bovine TLR9 protein sequence. T1 is the bovine TLR9 protein converted sequence data derived from this study while NM_183081.1 is *B. taurus* TLR9 protein converted sequence. Green colour shows the amino acid change caused by SNP 2238 and blue shows the amino acid change caused by SNP 2924.

However, SNPs are either "neutral" in the sense that the resulting point-mutated protein is not functionally different from the wild-type, or they are "non-neutral" in that the mutant and wild-type differ in function (Bromberg and Rost 2007). To find out if these SNPs were neutral or non-neutral, the protein sequence was entered into the software programme SNAP (https://rostlab.org/services/snap/) which evaluates the effects of single amino acid changes in protein function. The results were emailed back to me confirming the amino acid changes caused by SNPs (2238 and 2924) observed in all the *B. indicus* samples used in this study were neutral, implying that these amino acids are functionally the same as those observed in *B. taurus* RefSeq.

5.3.3.2 Frequency distribution of SNPs

SNP 2256 (rs55617244) was found in 22 out of 72 cattle samples (31.6%) while the remaining 50 (69.4%) did not have a SNP at that position. Seven out of the 22 SNPs (31.8%) were found in the Ugandan samples while the remaining 15 were found in samples from Nigeria towns; 3 from Tambes (13.6%), 6 each from Bokkos and Kachia (27.3% each). Of the 22 SNP 2256 found, 18 (81.8%) were heterozygotes while the remaining 4 (18.2%) SNPs were homozygotes (Table 5.3).

Table 5.3 SNPs frequency distribution

SNPs	Heterozygosity	Origins of cattle				Total
		Tambes	Bokkos	Kachia	Uganda	
SNP 2256	heterozygote	2	5	5	6	18
	homozygote	1	1	1	1	4
	N/A	16	12	13	9	50
Total		**19**	**18**	**19**	**16**	**72**
SNP 2538	heterozygote	5	10	6	3	24
	homozygote	13	6	13	13	45
	N/A	1	2	0	0	3
Total		**19**	**18**	**19**	**16**	**72**
SNP 2547	heterozygote	4	3	4	0	11
	homozygote	0	0	1	0	1
	N/A	15	15	14	16	60
Total		**19**	**18**	**19**	**16**	**72**
SNP 2583	heterozygote	4	7	4	3	18
	homozygote	2	3	1	0	6
	N/A	13	8	14	13	48
Total		**19**	**18**	**19**	**16**	**72**
SNP 2823	heterozygote	1	5	5	6	17
	homozygote	1	1	1	1	4
	N/A	17	12	13	9	51
Total		**19**	**18**	**19**	**16**	**72**
SNP 2865	heterozygote	1	3	5	6	15
	homozygote	1	3	1	1	6
	N/A	17	12	13	9	51
Total		**19**	**18**	**19**	**16**	**72**

SNP 2538 (rs209982925) was found in 69 out of the 72 samples (95.8%) while the remaining 3 (4.2%) did not have a SNP at that nucleotide position. Sixteen of the 69 SNPs (23.2%) were found in Ugandan samples and from the Nigerian towns; 18 (26.1%) were found in Tambes, 16 (23.2%) in Bokkos and 19 (27.5%) in Kachia. Also, 24 (34.8%) of the 69 were heterozygotes (Table 5.3).

SNP 2547 (rs518414196) was the most uncommon SNP found in samples with only 12 out of the total 72 samples (16.7%) having a SNP at that locus while the remaining 60 (83.3%) did not possess a SNP at the nucleotide position. The SNP was totally absent in all Ugandan samples investigated but among Nigerian samples 4 (33.3%) were found in Tambes, 3 (25%) in Bokkos and the majority 5 (41.7%) were found in the Kachia samples. Out of the 12 SNPs found, 11 (91.7%) were heterozygotes while the last 1 (8.3%) was a homozygote (Table 5.3).

SNP 2583 (rs55617221) was found in 24 out of 72 cattle samples (33.3%) while the remaining 48 (66.7%) had no SNP at the nucleotide position. Three of the 24 SNPs were found in the Ugandan bovine samples, while from the Nigerian samples 6 (25%) was found in Tambes, 10 (41.7%) in Bokkos and 5 (20.8%) in Kachia (Table 5.3).

SNP 2823 (rs55617235) and 2865 (rs55617220) both had the same proportion of SNPs. Twenty-one out the total 72 cattle samples (29.2%) were found to have SNPs at nucleotide position 2823 and 2865. They were equally distributed in the regions; 7 (33.3%) in Uganda, 2 (9.5%) in Tambes, 6 (28.6%) each in Bokkos and Kachia. However, the main observed difference was found in their heterozygosity. For SNP 2823, 17 (81%) of the SNPs were found to be heterozygotes while the remaining 4 (19%) were homozygotes and for SNP 2865 15 (71.4%) of the samples were heterozygotes while the remaining 6 (28.6%) were homozygotes (Table 5.3).

5.3.3.3 Correlation Analysis

A correlation analysis was performed on SPSS to measure the relationship between all SNPs found in cattle samples and thus identifying SNPs with a significant potential to predict the possession of another SNP. Our data did not meet the assumptions of using the common Pearson's correlation coefficient because our variables were categorical and not continuous, so the Kendall's Tau-b (nonparametric) correlation coefficient which makes no use of such assumptions was used to perform the correlation analysis (Bolboaca and Jäntschi 2006). Preliminary correlation analysis was performed against each of the six SNPs (SNP 2238, 2256, 2538, 2547, 2583, 2823 and 2865) using the three categorical responses (1= heterozygotes, 2= homozygotes, 3 = N/A if no SNP). Interestingly the result shows SNP2256 to be statistically correlated with SNP 2823 (P= <0.05; R= 0.964) and SNP 2865 (P = <0.05; R= 0.948). Also, SNP 2823 and SNP 2865 were shown to be statistically correlated to each other (P = <0.05; R = 0.984) and SNP 2538 was statistically correlated with SNP 2583 (P= <0.05; R = 0.474) (Table 5.4).

The R-value is the coefficient value; a value of +1 is a perfect positive association while a value of -1 is a perfect negative association. For all other arrangements the value lies between -1 and 1 and increasing values would imply increase in agreement between the rankings but where the rankings are independent then the R-value is zero (Bolboaca and Jäntschi 2006).

Table 5.4 SNPs correlations: Three coded responses.

Kendall's Tau_b Correlations							
SNPs	Measurements	SNP 2256	SNP 2538	SNP 2547	SNP 2583	SNP 2823	SNP 2865
SNP 2256	Correlation Coefficient	1.000	.040	-.202	.004	.964**	.948**
	Sig. (2-tailed)	.	.727	.080	.971	.000	.000
	N	72	72	72	72	72	72
SNP 2538	Correlation Coefficient	.040	1.000	.029	.474**	-.002	-.027
	Sig. (2-tailed)	.727	.	.803	.000	.983	.812
	N	72	72	72	72	72	72
SNP 2547	Correlation Coefficient	-.202	.029	1.000	.086	-.192	-.188
	Sig. (2-tailed)	.080	.803	.	.450	.097	.101
	N	72	72	72	72	72	72
SNP 2583	Correlation Coefficient	.004	.474**	.086	1.000	-.044	-.032
	Sig. (2-tailed)	.971	.000	.450	.	.695	.773
	N	72	72	72	72	72	72
SNP 2823	Correlation Coefficient	.964**	-.002	-.192	-.044	1.000	.984**
	Sig. (2-tailed)	.000	.983	.097	.695	.	.000
	N	72	72	72	72	72	72
SNP 2865	Correlation Coefficient	.948**	-.027	-.188	-.032	.984**	1.000
	Sig. (2-tailed)	.000	.812	.101	.773	.000	.
	N	72	72	72	72	72	72

**. Correlation is significant at the 0.01 level (2-tailed).

However, since the main purpose of performing a correlation analysis was to determine if possession of one SNP can predict the possession of another, categorical responses were re-coded in the SPSS package to “Yes” for where a cattle possess a SNP at a nucleotide position irrespective of their heterozygosity and to “No” where the cattle did not possess a SNP at that nucleotide position.

Table 5.5 SNPs correlations: Dichotomous responses.

Kendall's Tau_b Correlations							
SNPs		**SNP 2256**	**SNP 2538**	**SNP 2547**	**SNP 2583**	**SNP 2823**	**SNP 2865**
SNP 2256	Correlation Coefficient	1.000	.138	-.216	.043	.967**	.967**
	Sig. (2-tailed)	.	.244	.069	.719	.000	.000
	N	72	72	72	72	72	72
SNP 2538	Correlation Coefficient	.138	1.000	.093	-.147	.134	.134
	Sig. (2-tailed)	.244	.	.432	.214	.260	.260
	N	72	72	72	72	72	72
SNP 2547	Correlation Coefficient	-.216	.093	1.000	.079	-.205	-.205
	Sig. (2-tailed)	.069	.432	.	.505	.084	.084
	N	72	72	72	72	72	72
SNP 2583	Correlation Coefficient	.043	-.147	.079	1.000	.000	.000
	Sig. (2-tailed)	.719	.214	.505	.	1.000	1.000
	N	72	72	72	72	72	72
SNP 2823	Correlation Coefficient	.967**	.134	-.205	.000	1.000	1.000**
	Sig. (2-tailed)	.000	.260	.084	1.000	.	.
	N	72	72	72	72	72	72
SNP 2865	Correlation Coefficient	.967**	.134	-.205	.000	1.000**	1.000
	Sig. (2-tailed)	.000	.260	.084	1.000	.	.
	N	72	72	72	72	72	72

**. Correlation is significant at the 0.01 level (2-tailed).

Thus, the responses were recoded to be dichotomous and the Kendall's Tau_b (nonparametric) correlation coefficient analysis was performed using all 6 SNPs as before. Interestingly, correlation with the new data shows SNP 2538 and SNP 2583 to be statistically not correlated with each other (P = 0.214; R = 0.147). However, the coefficient value for the other SNP associations increased indicating strong positive association between them. SNP 2256 showed a stronger correlation with SNP 2823 (P = <0.05; R = 0.967) and SNP 2865 (P = 0.05; R = 0.967). Also SNP 2823 showed a perfect positive relationship with SNP 2865 (P = <0.05; R = 1) (Table 5.5).

5.3.3.4 Association of SNPs with trypanosome infection and origin of cattle

Trypanosome infection status data was available for only Ugandan samples. It was not possible to investigate the overall association of trypanosome infection status and relationship with these SNPs. However, association between trypanosome infection status and each SNP was tested but the results showed no association as P= >0.05 in all cases. Furthermore, analysis using SPSS showed that where cattle comes from is not statistically associated with possession of SNPs; SNP 2256 (P = 0.354), SNP 2538 (P = 0.399), SNP 2547 (P = 0.148), SNP 2583 (P = 0.126), SNP 2823 (P = 0.141) and SNP 2865 (P = 0.141).

5.3.4 Epigenetic analysis of variants in bovine TLR9 gene

Su et al. (2014) identified the bovine TLR9 gene to be methylated and describes DNA methylation as a major epigenetic marker playing a crucial role in transcriptional regulation, occurring at the 5' position of cytosine in CpG dinucleotides found mostly in CpG Islands. They are involved in inheritable alterations in gene expression that are not mediated by changes in DNA sequences and can be analysed using different approaches. However, no single method is appropriate for DNA methylation analysis and investigators are advised to use a method that suits their specific research needs (Shen and Waterland 2007). Since the bovine TLR9 gene was shown to have two CpG islands using EMBOSS Cpgplot (Fig. 5.8), our study was targeted at investigating if this gene is methylated and how it affects SNPs identified from DNA sequence data derived using a novel TLR9 hemi-nested PCR method. Analysis of bisulfite sequence data is based on chemical treatment of double-stranded DNA with sodium bisulfite which converts unmethylated cytosine into uracil, while methylated cytosine remains unchanged (Fig. 5.13). Subsequent PCR amplification converts all uracil to thymidine and the PCR products are then sequenced and compared against a reference sequence.

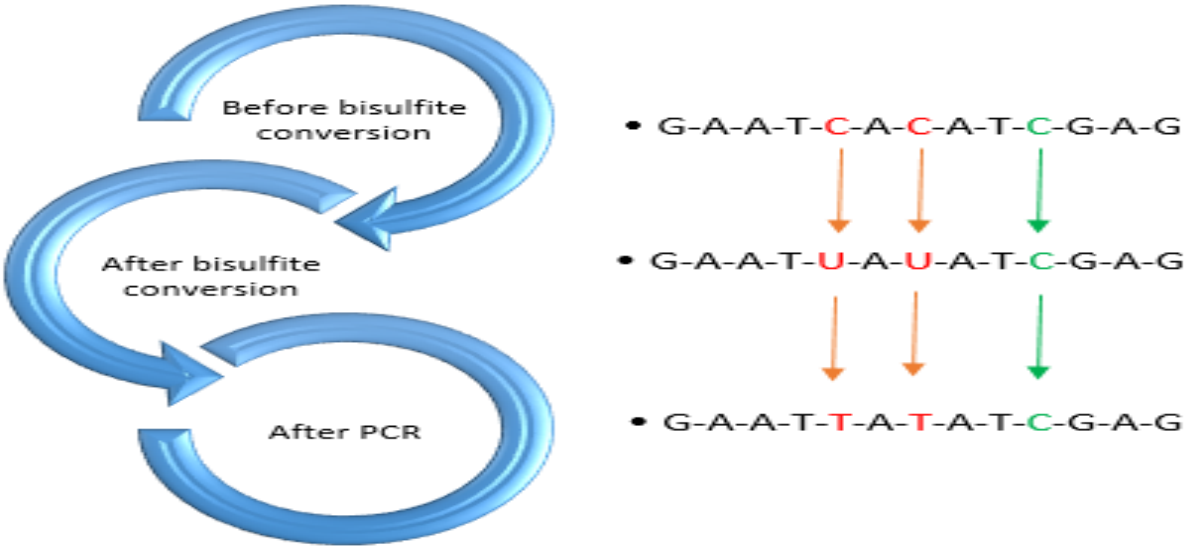


Figure 5.13 Detection of methylated Cytosines after bisulfite treatment. Blue arrows indicate the 3 major stages. Cytosine (red arrow) changes to U in stage 2 then to T after PCR if unmethylated. While methylated C's (green arrows) stay the same in all stages.

5.3.4.1 Development of bovine TLR9 bisulfite hemi-nested PCR

In order to commence bisulfite PCR, the first step was to identify if the gene has CpG islands since they are a target for DNA methylation. The EMBOSS Cpgplot used in this study identified two CpG islands which suggest a possibility that the gene is methylated. Having established this, two test bisulfite conversion kits (EZ DNA Methylation-Direct™ and EZ DNA Methylation-Gold™ Kit) were ordered from Zymo Research. Genomic bovine DNA (Amsbio bovine DNA) and DNA extracted from two of the study samples (U7 and U8) were converted to bisulfite DNA using these kits according to the manufacturer's recommendation. Bisulfite-conversion-based methylation primers targeting the two CpG islands were then designed using Methprimer (Li and Dahiya 2002) for novel amplification of methylated bovine TLR9 gene. The programme makes the assumption that all C's have been converted to T's, so the onus is for the investigator to select primers avoiding C-G sites (bisulfite PCR) and converting all C's to T's or include C-G sites in selecting primers (methylation-specific

PCR) and leave C's next to G's to stay the same. The C's next to G's stays the same if the gene is methylated even after bisulfite conversion (Fig. 5.14). The bisulfite PCR appears to be unbiased so was chosen as our preferred method for primer selection. To test whether the bovine DNA samples used had converted after bisulfite treatment and thus investigate DNA methylation on bovine TLR9 gene, two set of primers were selected for both CpG Island 1 and CpG Island 2. For CpG Island 1, the forward primers is bCpGF1 (5'-GGAAATTAGTTGAAGGTTTTGAGTAA-3') while the reverse primer is bCpGR1 (5'-CTATCTCTCCAACAAAAAATCCAC-3'), both primers would result in an estimated band size of 279bp then for CpG Island 2 the forward primer is bCpGF2 (5'-TTGGGATTTTTGGTATTGTTTTTAT-3') while the reverse primer is bCpGR2 (5'-AATTAACCCAAAAACTACCCTAACC-3'). The amplification of CpG Island using these primers would result in a band size of 500bp (Fig. 5.14).

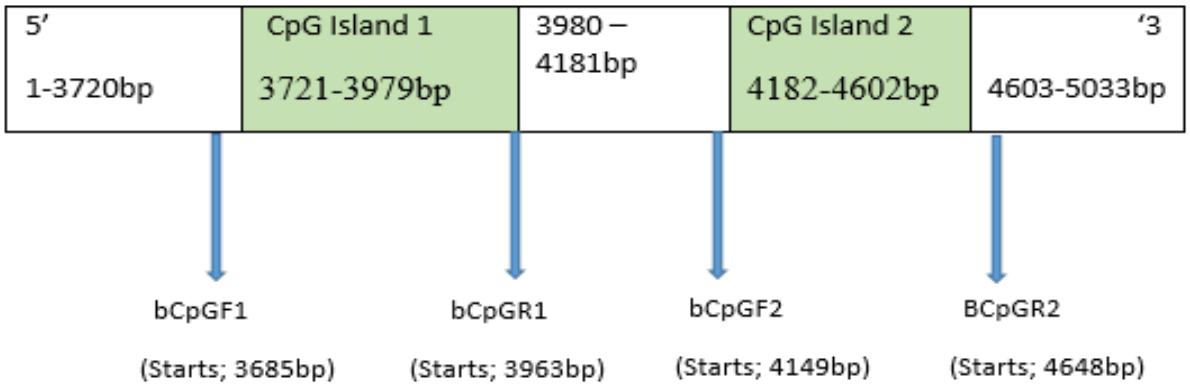


Figure 5.14 Bisulfite primer locations for bovine TLR9 gene. Arrows show primer locations on bovine TLR9 gene using the representative *B. taurus* TLR9 mRNA (Accession no: EF076731).

The EPIK amplification kit, a ready-to-use mixture for amplification of bisulfite converted DNA was purchased from Bioline and used for the PCR amplification according to the manufacturer's recommendation. The first set of bisulfite converted DNA tested were three samples (Amsbio bovine DNA, U7 and U8) treated using the EZ DNA Methylation-Direct™ kit. For each sample, both CpG Island 1 and CpG Island 2 were targets for amplification

using the primers described above. Unfortunately, the PCR amplification did not result in any visible bands after gel electrophoresis. Following several attempts of repeated experiments and optimisation there were still no bands present on the gel image indicating that either the bisulfite treatment had not worked or the primers are not working (Fig. 5.15).

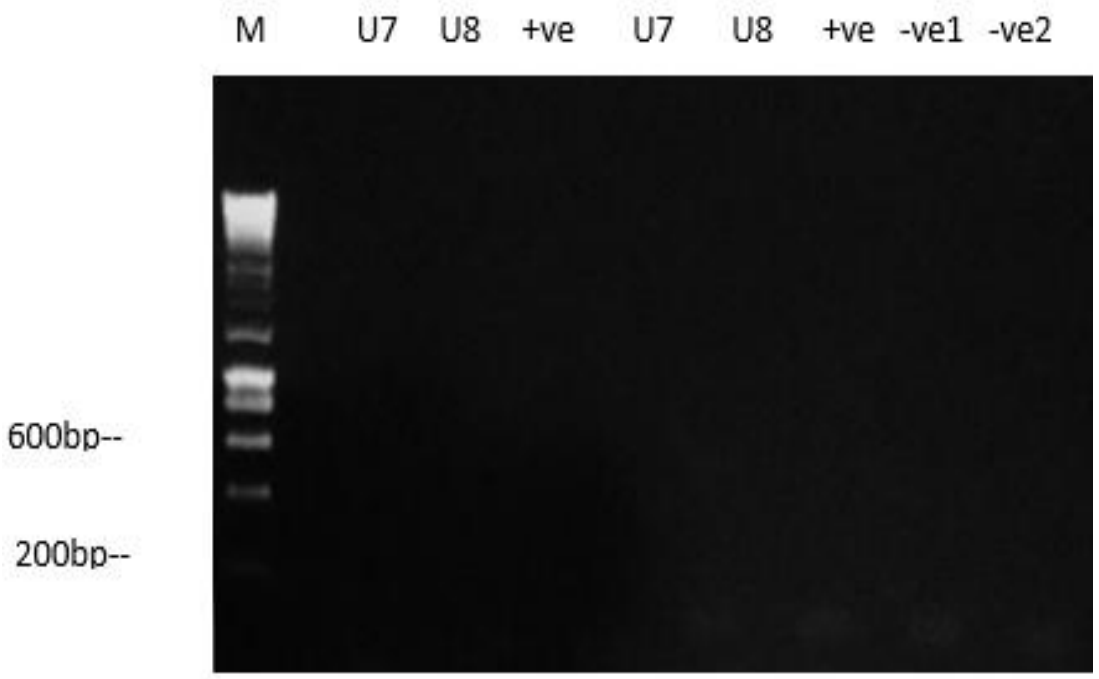


Figure 5.15 Failed bisulfite PCR using EZ methylation direct kit. Lane M is Hyperladder 1kb plus marker. –ve1 is untreated Amsbio bovine DNA, -ve2 is water and the other lanes are bovine samples treated with EZ methylation direct kit. The first three U7, U8 and +ve were amplified using the bisulfite CpG Island 1 primers were the others were amplified using bisulfite CpG Island 2 primers. No band on lanes indicate failed treatment or failed amplification.

Since the lack of visible bands could have been a result of failed bisulfite treatment, we tried treating another set of samples with the EZ DNA Methylation-Gold™ Kit and the reaction mixture consisted of the newly designed primers, EPIK amplification ready-to-go mixture from Bioline and water. Detailed procedure is described in Chapter two of this report. After bisulfite treatment and PCR amplification there were very faint bands shown in lanes on the gel, this gave us an indication the treatment had worked but probably the DNA concentration

is too low. 1ul each from all the samples were taking out and re-amplified using the same conditions as before. Interestingly, the results showed clear bands of two different sizes (279bp and 500bp) indicating that treatment and bisulfite-conversion-based primers have worked and both CpG Islands have been amplified (Fig. 5.16).

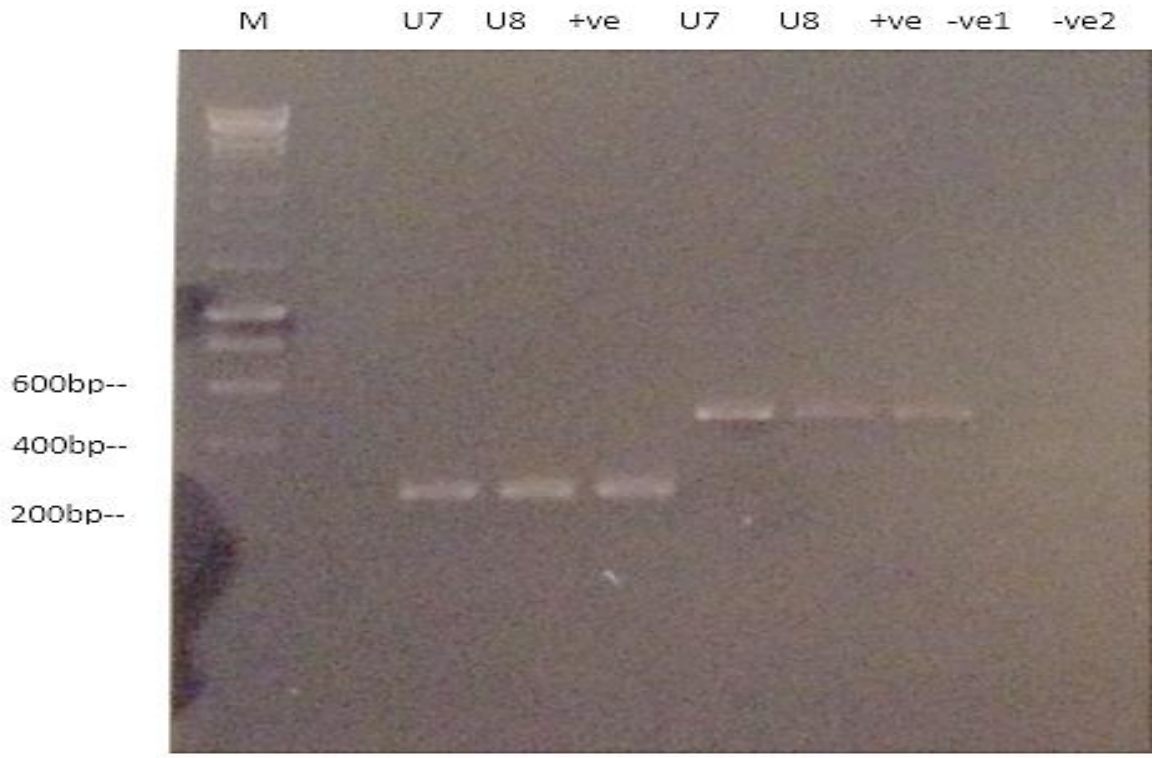


Figure 5.16 Bisulfite PCR using EZ methylation gold kit. Lane M is Hyperladder 1kb plus marker. –ve1 is untreated Amsbio bovine DNA, -ve2 is water and the other lanes are bovine samples treated with EZ methylation direct kit. The first three U7, U8 and +ve samples were amplified using the bisulfite CpG Island 1 primers while the others were amplified using bisulfite CpG Island 2 primers. Band sizes on gel indicate successful treatment and amplification.

Although, the bands shown above were not very strong they appeared due to re-amplification of PCR products using the same primers and the same conditions. This gave us the idea that if we applied a hemi-nested PCR approach we might have strong bands. The Hemi-nested bisulfite PCR was carried out by first amplifying the bisulfite treated DNA with the outer

primers bCpGF1 and bCpGR2 in the first round PCR, and using either bCpGF1 and bCpGR1 for amplification of CpG Island 1 (expected band size is 279bp) or bCpGF2 and bCpGR2 for amplification of CpG Island 2 (expected band size is 500bp) (Fig. 5.17). Seven samples were used to test this approach. The results showed stronger bands on the gel after amplification of both CpG Islands implying that the technique had worked (Fig. 5.18). CpG Island 1 and CpG Island 2 were amplified for a total of twenty samples. All products were sent to GATC for purification and sequencing. Detailed procedure for hemi-nested bisulfite TLR9 PCR is drafted in chapter two of this report.

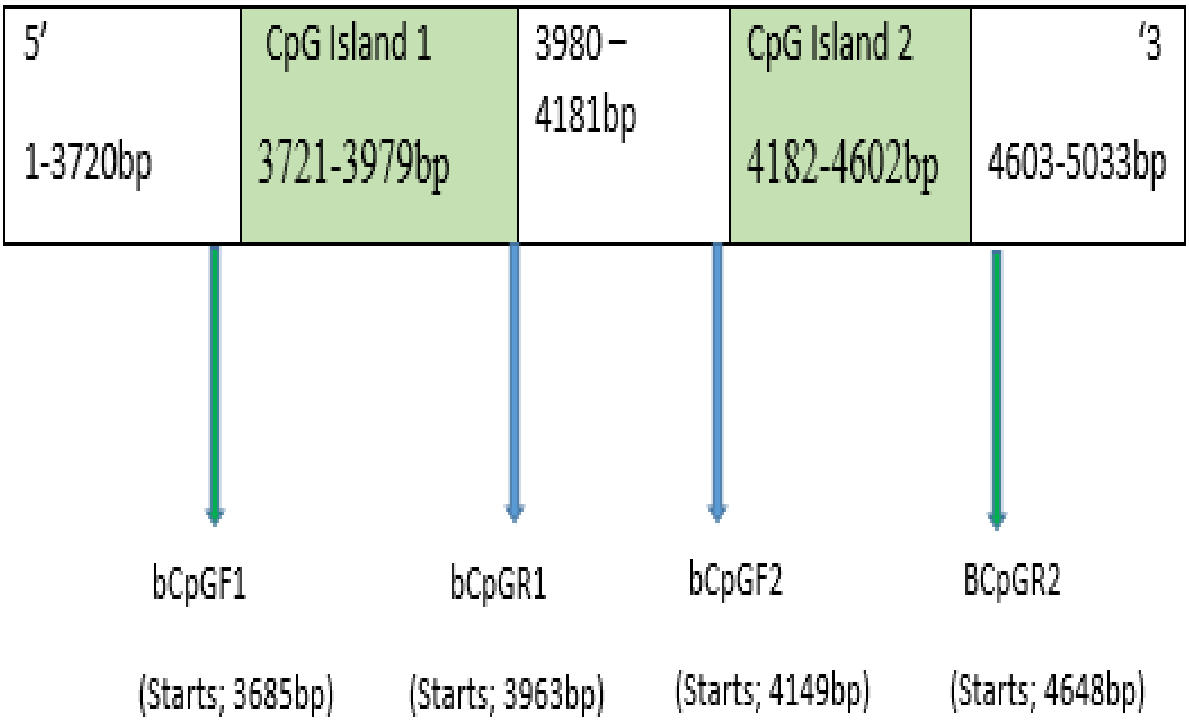


Figure 5.17 Hemi-Nested bisulfite PCR for bovine TLR9 gene. The green arrows are the outer primers used in 1st round PCR. The blue arrows are internal primers used in 2nd round to either amplify CpG Island 1 (bCpGF1 and bCpGR1) producing an estimated 279bp amplicon or CpG Island 2 (bCpGF1 and bCpGR1) producing an estimated 499 amplicon.

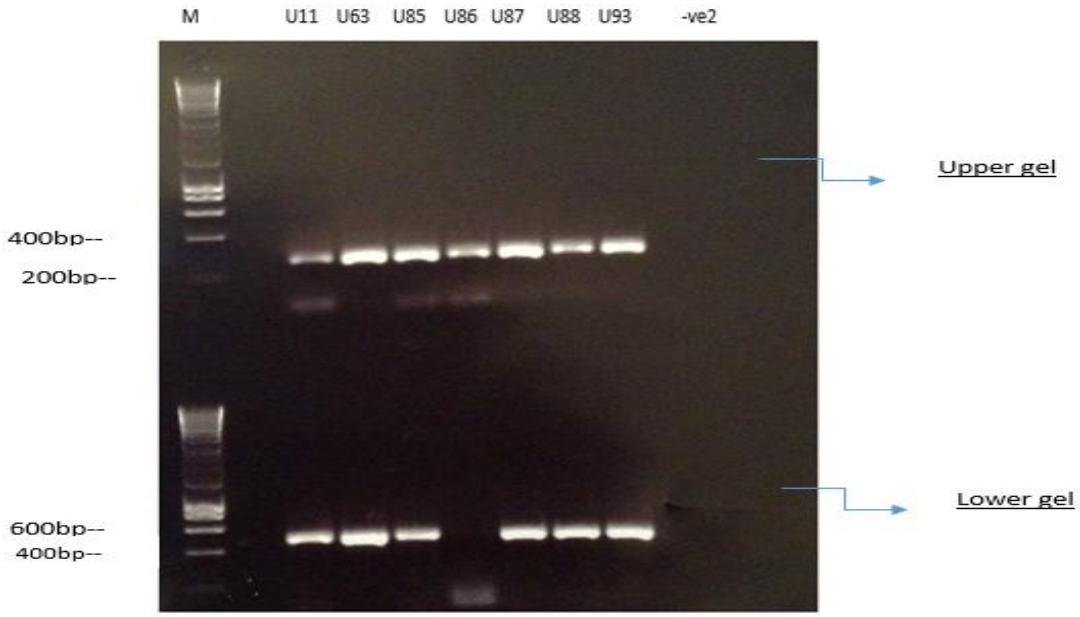


Figure 5.18 Successful Hemi-nested bisulfite PCR. The Upper gel shows amplification of CpG Island 1 (279bp) while the lower gel shows amplification of CpG Island 2 (500bp). Lane M is hyperladder 1kb plus marker, -ve is water while the other lanes are study samples. No band on U86 lower gel did due pipetting errors.

5.3.4.2 Bioinformatics and Statistical Analysis of TLR9 genetic variants

Bisulfite sequence data for all 20 bovine samples amplified using the newly developed bovine TLR9 bisulfite hemi-nested PCR were received electronically from GATC and analysed using Finch TV. The sequences were grouped into two, the first group consists of bisulfite sequence data for CpG Island 1 for each of the 20 samples while the second group consists of bisulfite sequence data for CpG Island 2 for each of the 20 samples. A multiple sequence alignment of each group was made using the cluster omega software (www.ebi.ac.uk/Tools/msa/clustalo/).

5.3.4.3 DNA Methylation analysis of bovine TLR9 CpG Island 1

After multiple sequence alignment of all bisulfite sequence data, the results showed that all C's had converted to T's. The only C's not converted were those next to a G implying that

the gene is methylated (Fig. 5.19). There were no alterations in any part of the whole alignment except for one observed at one locus, this was identified as SNP 2256.

```
U11          AGTAGTAATAGTATCGGTTTCGTGATTTTCGGTTTTTTCG      85
U86          AGTAGTAATAGTATCGGTTTCGTGATTTTCGGTTTTTTCG      92
u88          AGTAGTAATAGTATTGGTTTCGTGATTTTCGGTTTTTTCG      80
U9           AGTAGTAATAGTATCGGTTTCGTGATTTTCGGTTTTTTCG      92
U10          AGTAGTAATAGTATTGGTTTCGTGATTTTCGGTTTTTTCG      83
Amsbio       AGTAGTAATAGTATCGGTTTCGTGATTTTCGGTTTTTTCG      83
u93          AGTAGTAATAGTATCGGTTTCGTGATTTTCGGTTTTTTCG      83
u87          AGTAGTAATAGTATCGGTTTCGTGATTTTCGGTTTTTTCG      83
K12          AGTAGTAATAGTATTGGTTTCGTGATTTTCGGTTTTTTCG      83
K11          AGTAGTAATAGTATCGGTTTCGTGATTTTCGGTTTTTTCG      85
K15          AGTAGTAATAGTATCGGTTTCGTGATTTTCGGTTTTTTCG      84
Bisulfite    AGTAGTAATAGTATCGGTTTCGTGATTTTCGGTTTTTTCG      228
u61          AGTAGTAATAGTATCGGTTTCGTGATTTTCGGTTTTTTCG      80
K19          AGTAGTAATAGTATCGGTTTCGTGATTTTCGGTTTTTTCG      83
K13          AGTAGTAATAGTATCGGTTTCGTGATTTTCGGTTTTTTCG      84
U63          AGTAGTAATAGTATCGGTTTCGTGATTTTCGGTTTTTTCG      84
K18          AGTAGTAATAGTATCGGTTTCGTGATTTTCGGTTTTTTCG      85
u8           AGTAGTAATAGTATTGGTTTCGTGATTTTCGGTTTTTTCG      82
K17          AGTAGTAATAGTATCGGTTTCGTGATTTTCGGTTTTTTCG      86
K10          AGTAGTAATAGTATCGGTTTCGTGATTTTCGGTTTTTTCG      83
u7           AGTAGTAATAGTATCGGTTTCGTGATTTTCGGTTTTTTCG      84
K16          AGTAGTAATAGTATCGGTTTCGTGATTTTCGGTTTTTTCG      84
             ************** ************************
```

Figure 5.19 Multiple alignments of bisulfite CpG Island1 sequences. Only C's next to a G remained the same, implying the gene is methylated. Green colour shows the only alterations observed in the whole alignment. The variation was due to SNP 2256 (C/T) in that locus. Sample codes are on the left while nucleotide position is on t

Amongst all the SNPs identified in all cattle samples used in this study, SNP 2256 is the only SNP positioned in CpG Island 1, with a C/T allele. Interestingly, the SNP occurs at a methylation site because it is positioned next to G. So, normal genes not having the SNP would have a CpG at the loci which are sites for DNA methylation (e.g. U9, see Fig. 5.19). So when heterozygote, a sample has either a C/T binding to a G, so one allele has the CpG site while the other allele doesn't (e.g. U93, Fig. 5.19). However, when the sample is a homozygote at that locus then the CpG site is removed, as the sample would always have a TG at that locus (e.g. U88, Fig 5.19). Twenty-two cattle were found to have this SNP, 18 (81.8%) were heterozygotes while the remaining 4 (18.2%) SNPs were homozygotes. Having confirmed the scenario for CpG site alteration when a sample possesses this SNP, by

extrapolation we predicted the scenario for other samples that have SNP 2256 without the need for further bisulfite sequencing.

5.3.4.4 DNA Methylation analysis of bovine TLR9 CpG Island 2

Alignment of all sequences showed that all C's where converted to T's except for those C's next to G's in CpG sites, which again gives an indication that the bovine TLR9 gene is methylated. Analysis of the cluster alignment showed that there were no alterations in any part of the whole alignment except for two observed at two loci, these were identified to be due to SNP2823 and SNP2865. Amongst all SNPs found in the bovine samples used in this study, SNP 2823 and 2865 are the only two present in CpG Island 2. SNP 2823 has two alleles, either a C/G next to CpG site. So, for a sample without the SNP the original sequence after bisulfite sequencing is T-C-G because the first C would convert to a T (e.g. U63, Fig. 5.20). A heterozygote sample would have G/T-C-G (e.g. U93, Fig. 5.20) and a homozygote sample would have a G-C-G (e.g. U88, See Fig. 5.20). This implies than in any cases that have SNP 2823 does not result in removal of CpG sites in bovine TLR9 CpG Island 2. SNP 2823 was found in 21 out of 72 samples, of these, 17 were heterozygotes while 4 were homozygotes. By extrapolation, we predicted the scenario for the other samples that have SNP2823 without the need for further bisulfite sequencing.

Figure 5.20 Multiple alignments of bisulfite CpG Island2 sequences. Only C's next to a G remained the same, implying that the gene is methylated. Green colour show variants due to SNP 2823 while red colours shows those due to SNP2865. Numbers on right are nucleotide positions while codes on the left are sample codes.

Lastly, SNP 2865 was the second SNP for in the bovine TLR9 CpG Island 2. It has two alleles (A/G) occurring on the G bond of a CpG site. So a heterozygote sample normally has a C-A or C-G, but after bisulfite conversion the heterozygote samples were shown to have a T-A or T-G (e.g. U8 and U93, Fig. 5.20), while homozygote samples have a TA (U88, Fig. 5.20). This implies that whenever the cattle possess SNP2865, the C is always converted to T and, as a result, the CpG site is always removed which could impact on methylation. Twenty-one out of 72 samples sequenced were found to have SNP 2865, 15 were heterozygotes while 6 were homozygotes. By extrapolation, we predicted the scenario for the other samples that have SNP2865 without the need for further bisulfite sequencing.

5.3.4.5 Association of bovine TLR9 CpG SNPs with trypanosome infection

Trypanosome infection status data was available for only the 16 Ugandan samples. It was not possible to conclude on the overall association of trypanosome status and possession of any of these SNPs found in the CpG Islands. However, preliminary analysis using SPSS suggests

there is no association between trypanosome infection status and possession of CpG-SNPs as P = >0.05 in all cases. Nonetheless, all 3 SNPs are significantly correlated with each other (Table 5.5) and were found more frequently in Uganda than other regions of Nigeria.

5.3.5 Comparison of bovine TLR9 with human and mouse TLR9

The reference protein sequences for bovine TLR9 (NP_898904.1), human TLR9 protein (NP_059138.1) and mouse TLR9 protein (NP_112455.2) were identified on NCBI and a pairwise protein sequence alignment was performed using the EMBOSS Needle protein alignment software (http://www.ebi.ac.uk/Tools/psa/emboss_needle/). The purpose was to determine which of the two sequences (human or mouse) is more identical or similar to bovine TLR9 reference protein. The EMBOSS needle produces an identity percentage which is the percentage of identical matches between the two sequences over the region aligned including any gaps and it also produces a similarity percentage which is the percentage of matches between the two sequences over the region aligned including any gaps in length. The results showed that the human TLR9 protein is more similar and identical to the bovine TLR9 protein than to the mouse TLR9 protein (Table 5.6).

Table 5.6 Pairwise homology report for Bovine, human and mouse.

	Bovine TLR9 Protein	
	Similarity	Identity
Human TLR9 Protein	85%	79%
Mouse TLR9 Protein	81%	72%

Additionally, we used the EMBOSS Cpgplot to determine if either the human or mouse TLR9 DNA genomic reference sequences (NM_017442.3 and NM_031178.2) would also

possess CpG Islands which are sites of DNA methylation but neither of them possessed CpG Islands.

5.4 Discussion

This study reports the development of novel molecular approaches based on Hemi-nested PCR to investigate genetic and epigenetic variants of a specific locus in bovine TLR9 gene. A comparative genomic sequence analysis of 72 bovine samples showed a total of 9 SNPs in the target coding region of bovine TLR9 gene, of which 2 are nonsynonymous SNPs resulting in change in protein sequence while the other 7 SNPs are synonymous and do not result in change in the protein sequence. In previous study by Cargill and Womack (2007), 17 polymorphisms were found in bovine TLR9 gene, similarly 9 (2 nonsynonymous and 7 synonymous) were found in the coding region while the remaining SNPs were reported to have occurred in the intron or sequence flanking the 5' UTR. Unfortunately, a direct comparison cannot be made to ascertain if the SNPs and alleles found in that study are similar to those found in this study. This is due to the inability to identify the bovine TLR9 gene they used for comparison with their derived sequences. Nonetheless, this study takes a more appropriate approach as it compares the derived bovine TLR9 sequence data from all our samples against the *B. taurus* NCBI genomic reference sequence (NM_183081.1), this simplified the identification of SNPs in this study as any polymorphisms found were checked against bovine TLR9 SNPs already listed in the NCBI single nucleotide polymorphism database (NCBI dsSNP). Out of the 9 SNPs found in this study, 6 had been previously detected and referenced on NCBI. These are SNP 2256 (rs55617244), 2538 (rs209982925), 2547 (rs518414196), 2583 (rs55617221), 2823 (rs55617235) and 2865 (rs55617220). All of these were synonymous SNPs while, for the 3 remaining SNPs, 2 were nonsynonymous (SNP 2238 and SNP 2924) while the remaining one (SNP 2916) is synonymous. These three SNPs

were not referenced but were constant in all our samples and analysis using Cluster Omega to align our derived sequence with published bovine TLR9 sequences shows a 100% match but only two sequences (AJ509824.2 and AJ509825.1) and the RefSeq (NM_183081.1) was different from the rest of the sequences (See Appendix G). Details on NCBI shows the RefSeq was selected from the publication of the authors (Goldammer et al. 2004) of the other two sequences similar to it. Nonetheless, two of the SNPs (SNP 2238 and SNP 2924) led to amino acid changes, but these changes were shown to be neutral and so are functionally the same as those present in *B. taurus* (Fig. 5.12). A recent study by Lin et al. (2010) reported how 58 SNP markers were used to estimate the genetic diversity and structure between *B. taurus* and *B. indicus*. Therefore, this highlights the importance of identifying SNP markers between closely related species as they can be used to perform a genome-wide population genetic analysis. Also, there were variations in frequency of SNPs found in this study. SNP 2538 was the most frequent SNP found in the study population, 69 out of 72 of the samples (95.8%) had this SNP. All Ugandan samples possessed this SNP while the only 3 samples that did not have the SNP were from Nigeria. SNP 2547 was the least prevalent with only 12 out of the 72 samples (16.7%) having the SNP. However, none of the Ugandan samples appeared to have this SNP. Three particular SNPs (2256, 2823 and 2865) were of huge interest in this study because they were found in the CpG Islands. Interestingly, these SNPs were shown to occur more in Ugandan cattle (31.8%) than from cattle in any of the other 3 Nigerian villages, despite the Ugandan cattle constituting only 22.2% of the overall sample population (16 out of 72) (5.3). Unfortunately, literature on the analysis of variation in bovine TLR9 gene is still sparse so it was impossible to compare the prevalence of our findings with those from other studies. Nonetheless, the CpG Island SNPs (2256, 2823 and 2865) were shown to have a statistical significant correlation with each other after analysis using Kendall's Tau_b (Nonparametric) correlation coefficient analysis in SPSS. SNP 2256 was

significantly correlated with SNP 2823 (P = <0.05; R = 0.967) and SNP 2865 (P = <0.05; R = 0.967). Also SNP 2823 showed a perfect positive relationship with SNP 2865 (P = <0.05; R = 1) (Table 5.5). Thus, implying that the possession of one of these SNP (eg SNP 2256) can be used to predict the occurrence of the other two SNPs (i.e. SNP2823 and 2865). SNP 2256 is the only SNP in CpG Island 1 while SNP 2823 and 2865 are the only SNPs in CpG Island 2. Also the analysis of bisulfite sequence data from this study shows that the bovine TLR9 gene is methylated, as all C's converted to G's after PCR amplification except for C's positioned next to a G. The only alterations visualised after multiple alignments of our derived TLR9 sequences were due to SNP 2256 in CpG Island 1 and SNPs (2823 and 2865) in CpG Island 2. This study reports the conditions at which these CpG Island SNPs affect a C-G site which potentially could impact on methylation. For example, SNP 2256 occurs in a C-G site so cattle samples where homozygote for the SNP would have a T-G instead of the normal C-G and when heterozygote one allele is a T-G while the other is a C-G. SNP 2823, although occurs next to a CpG site does not occur directly on the site. Our analysis showed that in any scenario it does not remove a CpG site. However, having SNP 2865 guarantees a 100% chance of removing a CpG site. The SNP has two alleles (A/G) and occurs on the G bond of a CpG site so in any scenario (either homozygote or heterozygote) the C-G site is always removed and this could potentially have a major impact on methylation. The gaining or removal of CpG dinucleotides which are possible sites of DNA methylation could be a potential mechanism through which SNPs could affect the function of genes via epigenetics. Recent studies have shown the association of CpG-SNPs with diseases such as type 2 diabetes (Dayeh et al. 2013) and risky behaviours such as alcohol dependence (Taqi et al. 2011). Therefore, the identification of CpG-SNPs in bovine TLR9 gene could be a crucial epigenetic marker for diseases associated with bovine TLR9. Although previous studies have shown association between TLR9 gene and trypanosome infections (Drennan et al. 2005,

Harris et al. 2007), the trypanosome infection status for our samples was available only for the Ugandan samples so any analysis carried out may not give an estimate of the true association significance. However, by using the available data our analysis suggests there is no association between trypanosome infection status and where cattle comes from or the possession of any of the SNPs as P = >0.05 in all cases. Lastly, although bovine TLRs are described to be homologous to human TLRs (Medzhitov and Janeway 2000, Menzies and Ingham 2006), such homology could be rather functional and does not imply that their sequences are 100% identical. This study reports two pairwise protein sequence alignments using the EMBOSS needle, one between bovine and human TLR9 and the other between bovine TLR9 protein and mouse TLR9 protein (Table 5.3), which showed that the bovine TLR9 protein is more identical and similar to that of human than of mice. Similar confirmation of bovine TLR9 gene to be homologous to that of humans has also been reported previously (Griebel et al. 2005).

In conclusion, this study reports a novel molecular approach based on PCR that identified genetic and epigenetic variants of a candidate gene- bovine TLR9. It identifies SNP 2538 as the most prevalent in study sample. It also confirms the methylation of bovine TLR9 gene and identifies CpG-SNPs (2256 and 2865) which removes a C-G site and perhaps could alter DNA methylation and possibly gene expression, therefore identifying them as potential epigenetic markers for bovine TLR9 gene. Lastly, the study reports the significant correlation between CpG Island SNPs which suggests that possession of one of the Island SNPs is a predicting tool for possession of the others. A particular weakness in this study is the unavailability trypanosome infection status for all samples which perhaps would have led to a conclusion on association significance of our findings with the trypanosomes. Nonetheless, by extension the results of our findings could be applied to investigate other diseases associated with bovine TLR9 gene.

Chapter 6

Discussion and conclusion

This project reports the development of molecular tools for pan-*Trypanosoma* analysis and epigenetic of the host. The ITS-Nested PCR is one of such tools that have been shown to be effective in detecting species/subspecies of *Trypanosoma* in mammalian hosts (Cox et al. 2005, Cox et al. 2010, Wastling and Welburn 2011). The ITS-Nested PCR was therefore chosen as the preferred tool for investigating the prevalence of trypanosomes from two mammalian hosts: Southern Nigerian cattle (Chapter 3) and British badgers (Chapter 4). In order to investigate what impact infections with trypanosomes might have on a gene which plays a major role in defence within the host immune system, the bovine toll-like receptor (TLR) genes were investigated as a possible target. The bovine TLR genes, like other mammalian TLR genes play a crucial role in a host's innate immune system by recognising pathogenic signatures of invaders (Medzhitov and Janeway 2000, Menzies and Ingham 2006). The bovine TLR9 is such a distinguished member of the TLR genes that has been associated with clinically important African trypanosomiasis (Drennan et al. 2005, Harris et al. 2007). The bovine TLR9 gene was therefore chosen as the candidate gene for investigation (Chapter 5).

The ITS-Nested PCR successfully detected *Trypanosoma pestanai* in British badgers from Woodchester Park, Gloucestershire sampled in this study recording a 35.4% (25.9% - 46.2%; 95% CI) prevalence. Our recorded prevalence is not statistically different from the prevalence of 31% (25.6% - 37.1%; 95% CI) reported in a different study that surveyed badgers from Wytham Woods, Oxfordshire (Lizundia et al. 2011). The prevalences recorded from our study and the study by Lizundia et al. (2011) were higher than those from earlier studies (Macdonald et al. 1999, Pearce and Neal 1973) that were based on detecting trypanosomes

using traditional microscopical analysis. These studies reported lower prevalence rates of 10% and 7.7% respectively which suggests there is a higher degree of sensitivity when using molecular approaches. Although, the ITS-Nested PCR applied in this study appeared to have a similar sensitivity compared to the 18S PCR applied in the work of Lizundia et al. (2011), it has the added advantage of detecting trypanosomes from different mammalian hosts including mixed infections. A comparative analysis of 3 PCR-based methods for diagnosing trypanosomes showed that the ITS-Nested PCR detected more trypanosomes (28.1%) than the single ITS PCR (26.2%) and the species-specific PCR (10.7%) including mixed infections (Thumbi et al. 2008). The capability of the ITS-Nested PCR to detect mixed infection is one of its greatest strength that makes it unique compared to other molecular methods.

For example, by using the ITS-Nested PCR on Southern Nigerian cattle samples various species of trypanosomes (*T. Simiae*, 6.25%; *T. theileri*, 8.75%; *T.vivax*, 12.5%; unknown species, 5%; mixed infections 5%) were detected. The study (Chapter 3) also recorded a higher prevalence of 21.3% (13.6% - 31.5%; 95% CI) compared to previous studies (Abenga et al. 2004, Enwezor et al. 2009) that reported lower prevalence rates (9.1% and 8.4% respectively) which could be attributed to their approach of using traditional parasitological and immunological methods that offers lesser sensitivity compared to molecular approaches. Despite the high sensitivity of molecular-based approaches, the method of DNA extraction may also impact on the estimation of prevalence rate. For example, another study (Takeet et al. 2013) that used PCR-based technique to diagnose trypanosomes from Nigerian cattle reported a higher prevalence (63.7%) compared to the prevalence of 21.3% prevalence recorded in our study (Chapter 3). Their study was based on using DNA extracted from whole blood while our study used DNA prepared as punches taken out of blood spotted FTA cards. Using FTA cards as a sampling tool could result in lower than expected prevalence rates when diagnosing trypanosomes (Adams et al. 2008, Cox et al. 2010) which may be due

to the uneven distribution of the parasite around the FTA card matrix (Cox et al. 2010). Therefore, estimating an absolute prevalence using the FTA cards is dependent on the number of punches taken out of the card. Another challenge with the usage of FTA cards is that in some cases, even after detection of the parasite from a punch of FTA it could be impossible to fall back to the same FTA card to detect the same species of trypanosomes previously amplified. Such challenges can be overcome by applying the agarose gel-based touch preparation technique developed in this project (Chapter 4). By applying this method, we were able to isolate and preserve trypanosome DNA from bands shown on gels which could be more cost-effective compared to the usual purification kits. Unfortunately, just like these kits, the DNA preserved after the process can only be re-amplified using the same primers that produced the bands initially or re-amplified using new internal primers. So, this limits the researcher on what can be done with the available DNA (Chapter 3).

One limitation of the ITS-Nested PCR experienced in this project is its inability to always identify species of trypanosomes detected based on DNA sequence analysis of the PCR product, which is mostly, due to the absence of ITS genes of some/new species of trypanosomes in the databases. This makes it impossible to search the databases for identical sequences using DNA sequences derived from the new species amplified using ITS-Nested primers. However, with the rising influx of genes into the databases in recent times, such challenges could soon be overcome. An alternative approach for species identification after ITS-Nested PCR amplification is to apply an 18S PCR since the 18S genes for most trypanosomes are available on the databases. This could be useful particularly when the sample has been confirmed to be positive for trypanosomes using a more sensitive ITS-Nested PCR and the sample is also shown not to be infected with mixed species of trypanosomes. For example, one major challenge encountered in our study (Chapter 4) was identifying the species of trypanosomes that had been detected from the British badgers using

the ITS-Nested PCR. Initial sequencing of ITS-Nested PCR products produced DNA sequences showing partial alignment with *T. (Herpetosoma) otospermophili*, this was questionable as it is a known parasite of ground squirrels belonging to the *Herpetosoma* subgenus where host-specificity is a major criterion for naming species in the subgenus (Haore 1972). However, by developing and applying an 18S PCR we were able to identify the species of trypanosomes detected in the British badgers as *T. pestanai* (Chapter 4). This was achieved due to the abundant availability of 18S rRNA genes for most trypanosomes on the databases, which simplifies identifying of trypanosome species detected via PCR by comparing homology with published DNA sequences of known trypanosomes. The 18S PCR developed in our study is not species-specific so can be applied to the diagnosis of other trypanosomes. The highly conserved nature of the 18S rRNA gene makes it an important target for PCR and phylogenetic studies (Meyer et al. 2010). The 28S rRNA gene could be as important as the 18S rRNA gene because it is also highly conserved. It has been used as a target for molecular diagnosis of infectious diseases (Ahmad et al. 2002, Cnops et al. 2012, Sandhu et al. 1995) and phylogenetic studies (Hedges et al. 1990, Mallatt et al. 2004). However, studies conducted on molecular phylogeny or diagnostics of trypanosomes using the 28S rRNA gene as a target are still sparse and this could be due to the lack of 28S genes for most trypanosomes in the databases. The 28S sequence of *T. pestanai* is an example of an absent gene in the databases. However, the successful development of the 28S PCR and derivation of 28S alpha rRNA gene for *T. pestanai* in this study makes the gene available for molecular diagnostics and phylogenetic studies. For example, the 28S gene for *T. pestanai* derived in this study was included to form concatenated sequences (28S and 18S rRNA genes) to infer phylogenetic relationships between trypanosome species and compared with phylogenetic tree reconstructed using single genes (18S rRNA genes). Our results showed similarity between both trees, e.g. the badger trypanosomes derived from our study was

closely related to published *T. pestanai* sequences (100% bootstrap support) in both trees (Chapter 4). The 28S PCR technique could also be applied to derive 28S sequences of other absent trypanosome genes in the databases. One limitation, however, would be due to the unavailability of 28S sequences for most trypanosomes in the databases which could hinder comparison of homology between derived sequences from trypanosome positive sample with sequences published on the databases. The 28S PCR, therefore, would be more useful for deriving 28S sequences from trypanosome positive samples especially where the trypanosome species has already been identified, but the DNA sequence for the 28S region of the species is unavailable on the databases.

The project also reports the development and application of novel molecular approaches based on hemi-nested PCR to investigate genetic and epigenetic variants of specific loci in bovine TLR9 mRNA gene (Chapter 5). The target sequence for hemi-nested PCR amplification was the region covering two CpG Islands of the bovine TLR9 gene predicted using the Cpg EMBOSS plot software. The hemi-nested PCR is a modification of the nested-PCR technique and has also been applied in previous studies (Bachman et al. 2004, Djordjevic et al. 1998, Pujol-Rique et al. 1999). Our findings showed the technique to be highly sensitive and specific when applied suggesting it could be useful when the concentration of the DNA used as starting material is low for a single PCR and there is a lack of conserved regions of the gene/target sequence for primer selection. By applying the technique on the African bovine samples used for our study, good DNA sequence data was derived from each sample. Our sequences were compared with the reference sequence (RefSeq) for *B. taurus* TLR9 (NM_183081.1), there is currently no RefSeq entry for *B. indicus* TLR9 available on the databases. Following the analysis of the derived DNA sequence data, 9 polymorphisms (SNP 2238, 2256, 2538, 2547, 2583, 2823, 2865, 2916 and 2924) were observed in the sample population. Of these, 3 SNPs (2238, 2916 and 2924) were

constant in all of our samples. Our samples were of *B. indicus* origin while the RefSeq used for comparison is of *B. taurus* origin. Additionally, of the 3 constant SNPs, only the observed SNPs at loci 2238 and 2924 are non-synonymous causing amino acid changes from aspartic acid (D) to glutamic acid (E) and aspartic acid (D) to alanine (A) at amino acid position 705 and 934 respectively but these changes are neutral as they do not differ in function compared to those present in *B. taurus*. The other 6 SNPs are previously reported by different authors and listed on NCBI SNP database as SNP 2256 (rs55617244), 2538 (rs209982925), 2547 (rs518414196), 2583 (rs55617221), 2823 (rs55617235) and 2865 (rs55617220). SNP 2538 was the most frequent SNP found in the study population, 69 out of 72 of the samples (95.8%) had this SNP. Our nonparametric correlation analysis on all 6 SNPs using Kendall's Tau_b (nonparametric) correlation coefficient on SPSS showed three SNPs (2256, 2823 and 2865) to be statistically correlated with each other. SNP 2256 showed a strong correlation with SNP 2823 (P = <0.05; R = 0.967) and SNP 2865 (P = 0.05; R = 0.967). Also, SNP 2823 showed a perfect positive relationship with SNP 2865 (P = <0.05; R = 1) (Table 5.5). These 3 SNPs were also identified in the same sample population used in a previous study (Cargill and Womack 2007). On the NCBI SNP database these 3 SNPs were reported by just this same author who listed the same 9 accession numbers for each of the SNPs, these consist of both *B. taurus* and *B. indicus* breeds. However, after a cluster alignment of the published sequences we could not identify the SNPs so it was impossible to establish if the 3 SNPs were also correlated in the published sequences (See Appendix G). Understandably, each sequenced sample would normally have both a forward and a reverse sequence read. Most sequences published on NCBI nucleotide database represent just one of the sequence read (either the forward or reverse strand). It could be possible that the unpublished strand may have the SNPs information we required to carry out our analysis and make the comparison but these were not accessible. Interestingly, the 3 SNPs were also the only SNPs found in

both CpG Island 1 (SNP 2256) and CpG Island 2 (SNP 2823 and 2865). Also, these 3 SNPs were more common in Ugandan cattle (31.8%) samples than in the Nigerian samples (Bokkos, 27.2%, Kachia, 27.2%; Tambes, 13.6%) despite the Ugandan cattle constituting only 22.2% of the overall sample population (Chapter 5).

This project also reports the development of a novel bisulfite hemi-nested PCR applied to investigate epigenetic variation in bovine TLR9 gene. The bovine TLR9 gene was recently reported to be methylated in a study on bovine whole-genome methylation analysis (Su et al. 2014). Our analysis following a successful bisulfite sequencing after bisulfite conversion and amplification of the African bovine samples showed that the bovine TLR9 gene is methylated as all C's converted to T's except for C's bonding with a G. The only alterations observed on analysis of bisulfite sequence data from our samples were due to SNPs found in CpG Island 1 (SNP 2256) and CpG Island 2 (SNPs 2823 and 2865). The occurrence of SNPs in a CpG Island might have an impact on DNA methylation particularly when the SNPs occurs at a CpG site, as this might result in the gaining or removal of a CpG dinucleotide, which are possible sites for DNA methylation. For example, recent studies have shown association of CpG-SNPs with diseases such as type 2 diabetes (Dayeh et al. 2013) and risky behaviours such as alcohol dependence (Taqi et al. 2011). Our study reports the conditions at which the CpG Island SNPs observed in bovine TLR9 genes affect a C-G site which potentially could impact on methylation. Briefly, for SNP 2256, a homozygote cattle will have the C-G site removed and possess a T-G, heterozygote cattle would have one allele as a T-G while the other is a C-G. SNP 2823 does not occur directly on a C-G site so does not affect a site in any scenario. However, having SNP 2865 guarantees a 100% chance of removing a CpG site. The identification of these CpG SNPs with the potential to alter C-G sites in bovine TLR9 gene suggests they could be important epigenetic markers for investigating diseases associated with bovine TLR9. Trypanosome infection status for our samples was available only for the

Ugandan samples. Preliminary analysis using the available data suggests there is no association between trypanosome infection status and the possession of any SNPs as P = >0.05 in all cases. Based on the insufficient data this cannot give a conclusive estimate of our findings.

One limitation of our findings is the inability to interpret the exact role of methylation on bovine expression of TLRs/TLR9 gene. DNA methylation in mammals is tissue-specific and plays a crucial role in silencing of genes to prevent them from being expressed in wrong tissues (Khulan et al. 2006, Kitamura et al. 2007). However, literature in the role of DNA methylation in expression of TLRs, in general, is still sparse and there is still a gap in knowledge regarding the exact role of DNA methylation in expression of genes generally (Phillips 2008). One possible future avenue to explore would be to see if it is possible to quantitatively amplify RNA from FTA cards, this would offer an opportunity to investigate gene expression in methylated and unmethylated genes to see whether the methylation process switches the genes on or off. Nonetheless, our study identifies epigenetic markers which could be used as a platform to investigate infectious diseases associated with bovine TLR9. Although, there was no significant association between the bovine TLR9 polymorphisms observed in this study and trypanosomiasis (p=>0.05), the TLR9 gene is shown to be responsible for recognition of unmethylated bacterial CpG DNA (Hemmi et al. 2000) and polymorphisms in TLR9 has been associated with bovine pulmonary tuberculosis (Yang et al. 2013) and human *Mycobacterium tuberculosis* (Torres-Garcia et al. 2013). For future studies, the polymorphisms identified in our study could be useful epigenetic markers for investigating tuberculosis. Beyond the association of TLR9 polymorphisms with infectious diseases, SNPs have been useful to estimate genetic diversity and structure between *B. taurus* and *B. indicus* (Lin et al. 2010). Since the N'Dama cattle is believed to have a natural resistance to trypanosome infection (Courtin et al. 2008), Boran calves are

shown to be more susceptible (Paling et al. 1991) and TLR9 gene has been associated with African trypanosomiasis in experimental mice (Drennan et al. 2005, Harris et al. 2007), future studies may investigate if there are variations in TLR9 polymorphisms between the trypanotolerant N'Dama cattle and the susceptible Boran breed of cattle. Such studies might help understand the mystery behind resistance/susceptibility to the trypanosomiasis infection.

Another limitation of our study is that the sample size of 72 bovine samples used in Chapter 5 may not be enough for such type of studies aimed at investigating the polymorphisms in genes in relation to infectious diseases. One advantage, however, is that most of the SNPs identified from our samples have already been detected in other studies and there are no novel SNPs except for those SNPs (2238, 2916 and 2924) which were constant in all our samples after comparison with the NCBI *B. taurus* RefSeq for TLR9 (NM_183081.1). Also, for such studies, the use of next generation sequencing for samples would provide more quality DNA sequences for a more detailed analysis compared to the Sanger sequencing used in our study. For example, while the Sanger sequencing is restricted to producing DNA sequence reads of about 1,200 bp the next generation sequencing is a high-throughput method that can even be applied to whole-genome sequencing, identifying SNPs where necessary. It could be tailored to suit the user/consumer but by its application the entire bovine TLR9 mRNA could be easily analysed and every polymorphism identified for each individual sample. Commercial sequence companies such as Source Bioscience or GATC Biotech can provide this service.

In conclusion, this project shows that the ITS-Nested PCR works as a diagnostic tool for detection of trypanosomes in mammalian populations. It shows that both the novel 18S and 28S PCR developed in this project can also be used as molecular tools for diagnosing trypanosomes. Also, it shows that the novel hemi-nested PCR developed in the project can be applied to investigate polymorphisms in bovine TLR9 gene. It also shows the novel bisulfite

hemi-nested PCR developed in this project can be applied to identify and investigate epigenetic variants in bovine TLR9 mRNA gene. This project reports a 35.4% prevalence of *T. pestanai* infection in British badgers and the derivation of novel 28S alpha sequence data for the parasite. With comparable prevalence figures recorded from related studies (Lizundia et al. 2011), it suggests a significant proportion of UK badgers could be infected with trypanosomes. Also, the project reports an overall 21.3% (13.6% - 31.5%; 95% CI) prevalence of trypanosomiasis infection in Southern Nigerian cattle. It suggests that the higher prevalence recorded as compared to the lower prevalence from previous studies (Abenga et al. 2004, Enwezor et al. 2009) is due to high sensitivity and specificity of the ITS-Nested PCR that was used to screen the samples for this project. It identifies *T. vivax* as having the highest prevalence in the Southern Nigerian cattle sample population. The project reports the successful development of the agarose gel-based touch preparation which could be a cost-effective way of preserving trypanosome strains from bands shown on gels compared to the usual purification and preservation using kits. This project identifies 3 TLR9 SNPs (2238, 2916 and 2924) which were constant in all *B. indicus* samples in comparison with the NCBI RefSeq for *B. taurus* TLR9 gene but absent in all other published TLR9 sequences including those from our study. SNP 2538 was identified as the most prevalent in the study population. It applied bisulfite sequencing to confirm methylation of bovine TLR9 gene. The study identifies 2 CpG SNPs (2256 and 2865) capable of removing C-G sites, suggesting that this may have an impact on methylation and possibly the expression of the gene. The project also reports the correlation between all 3 CpG Island SNPs, implying the possession of each of these SNP can be used as a predicting tool for the possession of other two. Lastly, by extension some of the approaches and findings from the project could be useful in investigating other infectious diseases.

Appendices

Appendix A: DNA sequence alignment of badger trypanosomes with *T. otospermophili*

BADGER 12K ITS1 Alignments with *T. Otospermophili*

```
Query  1     CGGA-AGTCGGGTGTTGCACCTGACCGAAAGTTCACCGATATTTCTTCAATANAGGAAGC  59
             |||| |||||||||||| ||||||||||||||||||||||||||||||||||| |||||||
Sbjct  2134  CGGACAGTCGGGTGTTTCACCTGACCGAAAGTTCACCGATATTTCTTCAATAGAGGAAGC  2193

Query  60    AAAAGTCGTAACAAGGTAGCTGTAGGTGAACCTGCAGCTGGATCATTTTCCGATGA  115
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2194  AAAAGTCGTAACAAGGTAGCTGTAGGTGAACCTGCAGCTGGATCATTTTCCGATGA  2249
```

BADGER 12K ITS1 Alignments with *T. Otospermophili*

```
Query  1     GCTCAAGGAGGTACTCGGGTTTATGGCTCAATGGGCTAACACCTTCTTTGGCTCAATACT  60
             ||| ||||||| |||| ||||||||||||||||||||||||||||||||||||||||||
Sbjct  3621  GCTTAAGGAGGCACTCAGGTTTATGGCTCAATGGGCTAACACCTTCTTTGGCTCAATACT  3562

Query  61    CACATGAAGGGTTGTATTCAGCAACGGGTTTGCTTT-GCAATTGAATTTCGCCACTCCCG  119
             |||||||||||||||||||||||||||||||||| |||||||||||||||| ||| |
Sbjct  3561  CACATGAAGGGTTGTATTCAGCAACGGGTTTGCTTTTGCAATTGAATTTCGCCATTCCTG  3502

Query  120   CACCAGCCCTCGAAGCGACGCCTC-TGCCACAACTCAAGACCCTATAAGGCCCGATTCGG  178
             |||||||||||| |||||||||    ||||||||||||  ||||||||||||||||||||
Sbjct  3501  CACCAGCCCTCAAAGCGACGCTAATTGCCACAACTCAAAGCCCTATAAGGCCCGATTCGG  3442

Query  179   CAACGAGTTGGATCCCTCTTCGCTCGCCGCTGACTGAGAGAATCACGGTTGTTTTCTTTT  238
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3441  CAACGAGTTGGATCCCTCTTCGCTCGCCGCTGACTGAGAGAATCACGGTTGTTTTCTTTT  3382

Query  239   CCTCCGCTGAGTAATATGCTTAAGTTCGGCGGGTAGTCCTGCCACACTCAGGTCTGTAAA  298
             |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3381  CCTCCGCTTAGTAATATGCTTAAGTTCGGCGGGTAGTCCTGCCACACTCAGGTCTGTAAA  3322

Query  299   CAAAGTaaaaaa  310
             ||||  ||||||
Sbjct  3321  CAAA--AAAAAA  3312
```

BADGER 45P ITS1 Alignment Alignments with *T. Otospermophili*

```
Query  244   AAGCCCGGCCCCAACAACGGGTCGGGATGGATGACTTGGTTTCCTATTTCGTTGAAGAAC  303
             |||| |||||||||||||| |||| |||||||||||||| |||||||||||| |||||||
Sbjct  2648  AAGCACGGCCCCAACAACGTGTCGCGATGGATGACTTGGCTTCCTATTTCGTAGAAGAAC  2707

Query  304   GCAGCAAAGTGCGATAAGTGGTATCAATTGCAGAATCATTCAATTACCGAATCTTTGAAC  363
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2708  GCAGCAAAGTGCGATAAGTGGTATCAATTGCAGAATCATTCAATTACCGAATCTTTGAAC  2767

Query  364   GCAAACGGCGCATGGGAGAAGCTCTTTGGAGCCATCCCCGNGCATGCCATATTTCTCAGT  423
             |||||||||||||||||||||||||| ||| |||||||| |||||||||| ||||||||
Sbjct  2768  GCAAACGGCGCATGGGAGAAGCTCTTTTGAGTCATCCCCGTGCATGCCATAATTCTCAGT  2827

Query  424   GNCGAACAACaaaaaaaC  441
             | ||||||||||||||||
Sbjct  2828  GTCGAACAACAAAAAAAC  2845
```

BADGER 45P ITS2 Alignment Alignments with *T. Otospermophili*

```
Query  1     GCTGACTGAGAGAATCACGGTTGTTTTCTTTTCCTCCGCTGAGTAATATGCTTAAGTTCG  60
             |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Sbjct  3413  GCTGACTGAGAGAATCACGGTTGTTTTCTTTTCCTCCGCTTAGTAATATGCTTAAGTTCG  3354

Query  61    GCGGGTAGTCCTGCCACACTCAGGTCTGTAAACAAAGTaaaaaa  104
             ||||||||||||||||||||||||||||||||||  ||||||
Sbjct  3353  GCGGGTAGTCCTGCCACACTCAGGTCTGTAAACAAA--AAAAAA  3312
```

Appendix B: Alignment of badger trypanosome with *Trypanosoma pestanai*

```
Query  1     TATCCTATGGGCAATTCTGAATCCTACTGGGCAGCTTGGATCTCGTCCGTTGACGGAATC  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1788  TATCCTATGGGCAATTCTGAATCCTACTGGGCAGCTTGGATCTCGTCCGTTGACGGAATC  1729

Query  61    AACCAAACAAATCACTCCACCGACCAAAAGCGGCCATGCACCACCATTCAGGAAATCGAG  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1728  AACCAAACAAATCACTCCACCGACCAAAAGCGGCCATGCACCACCATTCAGGAAATCGAG  1669

Query  121   AAAGGACACTCAACCTGTCAATCCTTATCCTGTCCGGATCTGGTAAAGTTCCCCGTGTTG  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1668  AAAGGACACTCAACCTGTCAATCCTTATCCTGTCCGGATCTGGTAAAGTTCCCCGTGTTG  1609

Query  181   AGTCAAATTAAACCGCACGCTCCACGTCTTGTGGTGCCATTCCGTCAATTTCTTTAAGTT  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1608  AGTCAAATTAAACCGCACGCTCCACGTCTTGTGGTGCCATTCCGTCAATTTCTTTAAGTT  1549

Query  241   TCACTCTTGCGAACGTACTCCCCCCTGAGACTGTAACCTCAAAGCTTTCGCGTGAAGAAG  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1548  TCACTCTTGCGAACGTACTCCCCCCTGAGACTGTAACCTCAAAGCTTTCGCGTGAAGAAG  1489

Query  301   TCAGGTAACGTGCTGAGGATATTCCCGTCAAAGGCCCCCCACAATAAAGGAGGGACCTAA  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1488  TCAGGTAACGTGCTGAGGATATTCCCGTCAAAGGCCCCCCACAATAAAGGAGGGACCTAA  1429

Query  361   AAGAATATGCACGTAAATTTGATATTCATTGGCGAGGCAGGAAACGGAGAGGAGAGTGTC  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1428  AAGAATATGCACGTAAATTTGATATTCATTGGCGAGGCAGGAAACGGAGAGGAGAGTGTC  1369

Query  421   TGCGGGGCGAACCCTGACACACACGCCTCCCGAAATCTCCTGCCTACGACCAAAAACTCC  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1368  TGCGGGGCGAACCCTGACACACACGCCTCCCGAAATCTCCTGCCTACGACCAAAAACTCC  1309

Query  481   CCAATTCATGGGTGTCATCGTTTGCAGTGTGGACTACAATGGTCTCTAATCATCTTCGAT  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1308  CCAATTCATGGGTGTCATCGTTTGCAGTGTGGACTACAATGGTCTCTAATCATCTTCGAT  1249

Query  541   CCCCACACTTTGGTTCTTGATTGAGGAAGGTATCCTTGAAGAATGCCTTCGCTGTAGTTC  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1248  CCCCACACTTTGGTTCTTGATTGAGGAAGGTATCCTTGAAGAATGCCTTCGCTGTAGTTC  1189
```

```
Query  601   GTCCTGGTGCGGTCTAAGAATTTCACCTCTGACGCACCAATACGTTCTCCCCCGAACTAC  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1188  GTCCTGGTGCGGTCTAAGAATTTCACCTCTGACGCACCAATACGTTCTCCCCCGAACTAC  1129

Query  661   CCTCCTTCATTCCTGGATGCCGCGAGTCGAGAAAGAGGGGTGCAAGaaaaaaaatttttt  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1128  CCTCCTTCATTCCTGGATGCCGCGAGTCGAGAAAGAGGGGTGCAAGAAAAAAAATTTTTT  1069

Query  721   tGCACGGAACCTCCTCCTCGAAGGCGACGGGTGGCGCACTCCCCAAAAATGAGGTAGTGC  780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1068  TGCACGGAACCTCCTCCTCGAAGGCGACGGGTGGCGCACTCCCCAAAAATGAGGTAGTGC  1009

Query  781   CGCCGCACCCGACCGCTTGTCGCGCGGCCCCATAATCTCCAATGGACTTTTAAATACCAA  840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1008  CGCCGCACCCGACCGCTTGTCGCGCGGCCCCATAATCTCCAATGGACTTTTAAATACCAA  949

Query  841   CAAAAGCCGAAACGGTGGCCCATAGGCTGCTCCTTTGTTATCCCATGCTTTCTAATTCAA  900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  948   CAAAAGCCGAAACGGTGGCCCATAGGCTGCTCCTTTGTTATCCCATGCTTTCTAATTCAA  889

Query  901   GTCGACTGACTGCTTTGGTC  920
             ||||||||||||||||||||
Sbjct  888   GTCGACTGACTGCTTTGGTC  869
```

Appendix C: *Trypanosoma pestanai* 28S DNA sequence

AACCGTGATTCTCTCAGTCAGCGGCGAGCGAAGAGGGATCCAACTCGTTGCCGAATCGGGCCTTATAGGGTCT
TGAGTTGTGGCAGAGGCGTCGCTTCGAGGGCTGGTGCGGGAGTGGCGAAATTCAATTGCAAAGCAAACCCGTT
GCTGAATACAACCCTTCATGTGAGTATTGAGCCAAAGAAGGTGTTAGCCCATTGAGCCATAAACCCGAGTGCC
TCCTTGAGCGAACCAAGAGAGTAGCACTGCTTGGGAGTGCAGTGTCAAGCCGGCAGGTATGTGTCTGCCAAGG
TTAAATATAGAGTAGGAAGACCGATAGCGAACAAGTAGCGTGAGCGAAAGTTTGAAAAGTACTTTGGAGAGAG
AGTGACATAGAACCTGAAGTCGTGACAATGACAATGGAAGTACCTCCATTTCAACGGCCAAAACCAGCCTCAT
ACGGCAGCATCATCCCGCTGGGGTTAATTCTATCAATCCAGCGGGCCACGAGCGCAAACGATAGGGGCCATGG
TGAAACGGGCGGAGAGTATAAGGGGTTGAGAGGTTTCCTTTTTGGTCTCTCTCTTTCCCTTCTCCGTGATCCT
TGATGTTGGCGAAGGAAATGGGGTGCCCCCACCCGTCTTGAAACACGGACCAAGGAGTCAAACAGACGCGCAA
GGGAGAAGATATCGTTGCCAGTACTTCTCTCGTACTGAAAGGGAGATGCGAATGCATCTGTGGTGTGCACACA
CACAACATGATGAAGTAATGCCCAACACCGGCCGGCCCCGTTTTTGTGGGTTGAGTTGGAGTGTGCCTGTTTG
GACCCGAAAGGTGGTGAACTATGCCTGAACAAGGTGAAGCCCGGGGAAACCCAGGTGGAGGCCTGTAGCGATG
CTGACGTGCAAATCGCTCGCACGATTTGGGTATCGGAGCGAAAGACTCATCGAACCACCTAGTAGCTGGTTCA
CATCGAAGTTTCCCTCAGGATAGCTGGTGCTTGTTTGAAGAAGTATTGTGCGGTAAAGCAAATGATTAGAGGC
ATTGGTGTTCTTAGAATATCGACCTATTCTCAAACTTTAAATGTGCAAAGAAACCGTCCTTAGCTAGTATTTA
AGCAAGGGCGGACAGACAAGAATTAAGGCACCAAGTGGGCCTTCTTCGGTAAGCGGAGCAGGCGATGCGGAAT
GAACCGCTTAGGGATATGTGGCGTCAAAACTTGTGGGCTCATTCTCGATACGTGAAAAGGGTGTTGGTGGATA
GGGACAGTTGGACGGTGGCCATGGAAGTCGGCATCCGCTAAGGAGTGTGTAACAACTCACCAACCGAATCCAC
CGGCCCCGAAAATGGATGACGCTTAAGCCCAGTTAGTGATTGCCCATTATTCCTTTGGTAAGGGCGGAAAAGC
TTAAGAAAGCGTAGCCCTTCCGATGTAGATGTGGCGTGGAGGTAAGGACGAAGCTGATGGCGTGAGCCTCAGA
TGGACCAGCCTCTAGTGCAGATCTTGGTTGGCGTAGCAAAGATCTAACGGAGAATAAACTCAAAATGCAACGT
TGGATACTGGAGCGGGGAAGGATTTCGTGCCAACGGCACTCGTACACGAGTTGTTCGGATACTGAGCACAACG
TTACACCGTTTTGTTAGGAAAGTGAAGGTGCGTCGGCGGCATGCACGCCCCAGCGGTCTCTCCTTCTTTTATG
GGGGGGGAGGCCGTGTGCGTGCCGCTGGCCATGACTGAAAAGGGGCAACAGAGAACCTGGGATTATATTCCCA
AAAAGAATTTGCATGTGGGA

Appendix D: Alignment of 18S and Badger 18S trypanosomes DNA sequences

All sequences were aligned with Muscle and trimmed using Phylogeny.fr webserver before being used construct phylogenies (See Chapter 4)

```
>gi|295385|gb|L14840.1|TRYRRSS_Trypanoplasma_borreli_Pg-
JH_ribosomal_RNA_small_s
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTCCGACCTGAATTACCAAGCATGGGA
TAACA
AAGCAGCAACCTACGGTACACCGTTATGGCTTTTGTTGGTTTTAATAGTCTGTGGGACATTGTAGGAATGTGTGA
CGGAA
TGAAGGAGAGTTGTCGGGGGAGAACGTACTTGAACGTGAGAGGTGAAATTCTTAGACCGCTCAAAGACGAACTAC
AGCGA
AGGCATTCTCCAAGGATACTTTCCTCAATCAAGAACCAAAGTGTGGAGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACCACAAACGATGACACCCATGAATTGGGGTTTTTTAATTTACGTGCATTTTCAACTACGGGAATATCCTCAGCA
CGTTT
TTTACGCGAAAGCTTGAGTTATAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCTTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGAACCAGGATTGTCCAAAAGACAGCAATCTTTGGAAGCTACCTC
C

>gi|4468765|emb|AJ009159.1|_Trypanosoma_pestanai_18S_rRNA_gene,_isolate_LEM
_110
GGGCGTCCGTGATTTTTACTGTGACTAAAAAAGTGCGACCAAAGCAGTCGTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATGGGCCACCGTTTCGGCTTTTGTTGGTTTTAAAAGTCCATTGGAGATTATGGGGCCGCGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTATTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAGGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA
CCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGATAAGGATT
GACAG
GTTGAGTGTCCTTTCTCGATTTCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCGATCCTCGGGGGCGATATT
C

>Badger_tryp_18S
GGGCGTCCGTGATTTTTACTGTGACTAAAAAAGTGCGACCAAAGCAGTCGTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATGGGCCACCGTTTCGGCTTTTGTTGGTTTTAAAAGTCCATTGGAGATTATGGGGCCGCGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTATTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAGGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA
CCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
```

CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGATAAGGATT
GACAG
GTTGAGTGTCCTTTCTCGATTTCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCGATCCTCGGGGGCGATATT
C

>gi|148529690|gb|EF375882.1|_Trypanosoma_sp._carpio_18S_ribosomal_RNA_gene,
_part
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTCCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATGAGCTACCGTTTCGGCTTTTGTTGGTTTTAAAAACTCATTGGAGATTATGGAGTTGTGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
GGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTT
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATT
C

>gi|160878089|gb|EU185634.1|_Trypanosoma_ophiocephali_18S_small_subunit_rib
osoma
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTCCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATGAGCTACCGTTTCGGCTTTTGTTGGTTTTAAAAACTCATTGGAGATTATGGAGTTGTGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTT
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATT
C

>gi|4468767|emb|AJ009161.1|_Trypanosoma_rotatorium_18S_rRNA_gene,_isolate_B
2-II
GGGCGTCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTCCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATGGGCCACCGTTTCGGCTTTTGTTGGTTTTAAACGTCCATTGGAGATTATGGGGCTGTGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACTAAAGTATGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAT
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTGTTACGGGAATATCCTCAGCA
CGTTT

TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGGATTGTCCATAGGATAGCAATCCCTGGGGGCGGTATC
C

>gi|2808546|emb|AJ223567.1|_Trypanosoma_mega,_small_subunit_ribosomal_RNA_g
ene
GGGCGTCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTCCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATGGGCCACCGTTTCGGCTTTTGTTGGTTTTAAAAGTCCATTGGAGATTATGGGGCTGTGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTATAACGGGAATATCCTCAGCA
CGTTT
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGGATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATC
C

>gi|4468766|emb|AJ009160.1|_Trypanosoma_rangeli_18S_rRNA_gene,_isolate_RGB_
(Base
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATGGGCCACCGTTTCGGCTTTTGTTGGTTTTAAAAGTCCATTGGAGATTATGGGGCAGTGTGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTAACGGGAATATCCTCAGCA
CGTTA
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATT
C

>gi|320443936|gb|FJ001666.2|_Trypanosoma_dionisii_culture-
collection_TCC/USP:211
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATGGGCCACCGTTTCGGCTTTTGTTGGTTTTAAAAGTCCATTGGAGATTATGGGGCAGTGTGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC

ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATT
C

>gi|320443935|gb|FJ001665.2|_Trypanosoma_cruzi_marinkellei_culture-
collection_TC
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATGGGCCACCGTTTCGGCTTTTGTTGGTTTTAAAAGTCCATTGGAGATTATGGGGCAGTGTGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATT
C

>gi|4468746|emb|AJ009147.1|_Trypanosoma_cruzi_18S_rRNA_gene,_isolate_Sylvio
_X10
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTCTGGGCCACCGTTTCGGCTTTTGTTGGTTTTAAAAGTCCATTGGAGATTATGGGGCAGTGTGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATT
C

>gi|4468773|emb|AJ009163.1|_Trypanosoma_theileri_18S_rRNA_gene,_isolate_TRE
U_124
GGGCGTCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATGGGCCACCGTTTCGGCTTTTGTTGGTTTTAAAAGTCCATTGGAGATTATGGCGTCGTGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGCGCGTCAGAGGTGAAATTCTTAGACCGCGCCAAGACGAACTAC
AGCGA

AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAATTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCCAGGGGCGATATT
C

>gi|301793082|emb|FN666409.1|_Trypanosoma_melophagium_partial_18S_rRNA_gene
,_iso
GGGCGTCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATGGGCCACCGTTTCGGCTTTTGTTGGTTTTAAAAGTCCATTGGAGATTATGGCGTCGTGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGCGCGTCAGAGGTGAAATTCTTAGACCGCGCCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAATTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCCAGGGGCGATATT
C

>gi|46091659:375-
2593_Trypanosoma_grosi_genes_for_IGS,_18S_rRNA,_5.8S_rRNA,_28S
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATAGGCCACCGTTTCGGCTTTTGTTGGTTTTAACAGTCTATTGGAGATTATGGGGCTGTGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATT
C

>gi|2808495|emb|AJ223566.1|_Trypanosoma_lewisi,_small_subunit_ribosomal_RNA
_gene
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATAGGCCACCGTTTCGGCTTTTGTTGGTTTTAACAGTCTATTGGAGATTATGGGGCTGTGCGA
CGGAA

TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATT
C

>gi|2808547|emb|AJ223568.1|_Trypanosoma_musculi,_small_subunit_ribosomal_RN
A_gen
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATAGGCCACCGTTTCGGCTTTTGTTGGTTTTAACAGTCTATTGGAGATTATGGGGCTGTGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATT
C

>gi|4468759|emb|AJ009158.1|_Trypanosoma_microti_18S_rRNA_gene,_isolate_TRL_
132
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATAGGCCACCGTTTCGGCTTTTGTTGGTTTTAACAGTCTATTGGAGATTATGGGGCTGTGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATT
C

>gi|46091661:28-
2242_Trypanosoma_otospermophili_genes_for_IGS,_18S_rRNA,_5.8S_rR
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTTCGACTTGAATTAGAAAGCATGGGA
TAACA

AAGGAGCAGCCTATAGGCCACCGTTTCGGCTTTTGTTGGTTTTAACAGTCTATTGGAGATTATGGGGCTGTGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATT
C

>gi|46091662:28-
2245_Trypanosoma_kuseli_genes_for_IGS,_18S_rRNA,_5.8S_rRNA,_28S
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATAGGCCACCGTTTCGGCTTTTGTTGGTTTTAACAGTCTATTGGAGATTATGGGGCTGTGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATT
C

>gi|4468741|emb|AJ009142.1|_Trypanosoma_brucei_rhodesiense_18S_rRNA_gene,_i
solat
GGGCGTCCGTGAATTTTACTGTGACCAAAAAAGTGCGACCAAAGCAGTCGCCGACTTGAATTACAAAGCATGGGA
TAACG
AAGCATCAGCCCTGGGGCCACCGTTTCGGCTTTTGTTGGTTTTAGAAGTCCTTGGGAGATTATGGGGCCGCGTGC
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCAAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACCATGACACCCATGAATTGGGGAACATCAATTTACGTGCCTATTCTATCACGGGAATATCCTCAGCA
CGTTT
CTTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATGGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGTGCCGGGATTGTCCACAGGACAGCAGTCCCTGTGGTCGTCATC
C

>gi|4468740|emb|AJ009141.1|_Trypanosoma_brucei_gambiense_18S_rRNA_gene,_iso
late

GGGCGTCCGTGAATTTTACTGTGACCAAAAAAGTGCGACCAAAGCAGTCGCCGACTTGAATTACAAAGCATGGGA
TAACG
AAGCATCAGCCCTGGGGCCACCGTTTCGGCTTTTGTTGGTTTTAGAAGTCCTTGGGAGATTATGGGGCCGCGTGC
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCAAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACCATGACACCCATGAATTGGGGAACATCAATTTACGTGCCTATTCTATCACGGGAATATCCTCAGCA
CGTTT
CTTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATGGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGTGCCGGGATTGTCCACAGGACAGCAGTCCCTGTGGTCGTCATC
C

>gi|4468753|emb|AJ009154.1|_Trypanosoma_evansi_18S_rRNA_gene,_isolate_E_110
GGGCGTCCGTGAATTTTACTGTGACCAAAAAAGTGCGACCAAAGCAGTCGCCGACTTGAATTACAAAGCATGGGA
TAACG
AAGCATCAGCCCTGGGGCCACCGTTTCGGCTTTTGTTGGTTTTAGAAGTCCTTGGGAGATTATGGGGCCGCGTGC
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCAAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACCATGACACCCATGAATTGGGGAACATCAATTTACGTGCCTATTCTATCACGGGAATATCCTCAGCA
CGTTT
CTTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATGGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGTGCCGGGATTGTCCACAGGACAGCAGTCCCTGTGGTCGTCATC
C

>gi|4468768|emb|AJ009162.1|_Trypanosoma_simiae_18S_rRNA_gene,_isolate_KEN_2
GGGCGTCCGTGATCTTTACTGTGACTAAAAAAGAGCGACCAAAGCAGTCCACGACTTGAATTACAAAGCATGGGA
TAACA
AAGCATCAGCCCCAGGGCCACCGTTTCGGCTTTTGTTGGTTTTAGAAGCCCAGGGGAGATTATGGCGCCGCGTGC
CGGAA
TGAATGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGTATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACCATGACACCCATGAATTGGGGACCATTAACTTACGTGCCTATTCTGCCACGGGAATATCCTCAGCA
CGTTG
TTTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATGGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGGCCCGTGATCGTCCACAGGACAGCATTCCGTCGGAGCACTATC
A

>gi|4468743|emb|AJ009144.1|_Trypanosoma_congolense_(kilifi)_18S_rRNA_gene,_
isola

GGGCGTCCGTGATTTTTACTGTGACTAAAAAAGTGCGACCAAAGCAGTCGCCGACTTGAATTACAAAGCATGGGA
TAACA
AAGCACCAGCCCTAGGGCCACCGTTTCGGCTTTTGTTGGTTTTAGAAGTCCTTGGGAGATTATGGTGTCGCGTGC
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCAAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACCATGACACCCATGAATTGGGGACCACCAACTTACGTGTCTATTCTGCTACGGGAATATCCTCAGCA
CGTTG
TTTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGGCCCGTGATCGTCCACAGGACAGCCTACCGTGGCGGCGCTATC
A

>gi|4468745|emb|AJ009146.1|_Trypanosoma_congolense_(savannah)_18S_rRNA_gene
,_iso
GGGCGTCCGTGATTTTTACTGTGACTAAAAAAGTGCGACCAAAGCAGTCGCCGACTTGAATTACAAAGCATGGGA
TAACA
AAGCACCAGCCCTAGGGCCACCGTTTCGGCTTTTGTTGGTTTTAGAAGTCCTTGGGAGATTATGGCCTCGCGTGC
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCAAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACCATGACACCCATGAATTGGGGACCACCAACTTACGTGTCTATTCTGCTACGGGAATATCCTCAGCA
CGTTG
TTTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGGCCCGTGATTGTCCACAGGACAGCCTACCGTGGCGGCGCTATC
A

>gi|4468744|emb|AJ009145.1|_Trypanosoma_congolense_(riverine_forest)_18S_rR
NA_ge
GGGCGTCCGTGATTTTTACTGTGACTAAAAAAGTGCGACCAAAGCAGTCGCCGACTTGAATTACAAAGCATGGGA
TAACA
AAGCACCAGCCCTGGGGCCACCGTTTCGGCTTTTGTTGGTTTTAGAAGTCCTTGGGAGATTATGGCGCCGCGTGC
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCAAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACCATGACACCCATGAATTGGGGACCACCAACTTACGTGTCTATTCTGCTACGGGAATATCCTCAGCA
CGTTG
TTTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGGCCCGTGATTGTCCACAGGACAGCCTACCGTGGCGGCGCTATC
A

Appendix E: Alignment of 28S, 18S and badger trypanosomes DNA sequences

All sequences were aligned with Muscle and trimmed using Phylogeny.fr webserver before being used construct phylogenies (See Chapter 4)

>gi|295385|gb|L14840.1|TRYRRSS_Trypanoplasma_borreli_Pg-JH_ribosomal_RNA_small_s
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTCCGACCTGAATTACCAAGCATGGGA
TAACA
AAGCAGCAACCTACGGTACACCGTTATGGCTTTTGTTGGTTTTAATAGTCTGTGGGACATTGTAGGAATGTGTGA
CGGAA
TGAAGGAGAGTTGTCGGGGGAGAACGTACTTGAACGTGAGAGGTGAAATTCTTAGACCGCTCAAAGACGAACTAC
AGCGA
AGGCATTCTCCAAGGATACTTTCCTCAATCAAGAACCAAAGTGTGGAGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACCACAAACGATGACACCCATGAATTGGGGTTTTTTAATTTACGTGCATTTTCAACTACGGGAATATCCTCAGCA
CGTTT
TTTACGCGAAAGCTTGAGTTATAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCTTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGAACCAGGATTGTCCAAAAGACAGCAATCTTTGGAAGCTACCTC
C
AGAGTAGGAAGACCGATAGCGAACAAGTAGCGTGAGCGAAAGTTTGAAAAGCACTTTGGAGAGAGAGTGACGTAG
AACCT
GAAGTCGTGATAATGATAATGAAAGCACCTCCATTTCAGTGGCCAAAACCAGCCCTACATATCCAGGAGGTTGAC
TTTCT
TCTGGTATTGGTGAAGTTCATTCCCGTGTGCACAATGTTGGCGAAGGAAATGGGGTGTCTTCACCCGTCTTGAAA
CACGG
ACCAAGGAGTCAAACAGACGCGCAATTGCCAGTACTTCTCTCGTACTGAAAGGGAGCTGCCTACGCAGCCATGGT
GTGCA
CGCCCAACACCGGCCGGCCCCGGCTGAGCTTGAGTGTGCCTGTTTGGACCCGAAAGGTGGTGAACTATGCCTGAA
CAAGG
TGAAGCCGGGGGAAACCCAGGTGGAGGCCTGTAGCGATGCTGACGTGCAAATCGCTCGTATGATTTGGGTATCGG
AGCGA
AAGACTCATCGAACCACCTAGTAGCTGGTTCACATCGAAGTTTCCCTCAGGATAGCTGGTGCTTGTCACGGTAAA
GCAAA
TGATTAGAGGCTTTGGTTCTCTTAGAGTATCGACCTATTCTCAAACTTTAAATGTGTAGAAAAACTGTCAAGATC
AGAAG
AAAAGAATTAAGGCACCAAGTGGGCCTTCTTCGGTAAGCGGAGCAGGCGATGCGGAATGAACCGCTTAGGGATAT
GTGGC
GTCAAAACTTATGGGCTCATTCTGATACGTGAAAAGGGTGTTGGTGGATGAGGACAATTGGACGGTGGCCATGGA
AGTCG
GCATCCGCTAAGGAATGTGTAACAACTCACCAACCGAATCCACCGG

>gi|4468765|emb|AJ009159.1|_Trypanosoma_pestanai_18S_rRNA_gene,_isolate_LEM_110
GGGCGTCCGTGATTTTTACTGTGACTAAAAAAGTGCGACCAAAGCAGTCGTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATGGGCCACCGTTTCGGCTTTTGTTGGTTTTAAAAGTCCATTGGAGATTATGGGGCCGCGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTATTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAGGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC

ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA
CCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGATAAGGATT
GACAG
GTTGAGTGTCCTTTCTCGATTTCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCGATCCTCGGGGGCGATATT
C

>Badger_tryp_18S
GGGCGTCCGTGATTTTTACTGTGACTAAAAAAGTGCGACCAAAGCAGTCGTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATGGGCCACCGTTTCGGCTTTTGTTGGTTTTAAAAGTCCATTGGAGATTATGGGGCCGCGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTATTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAGGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA
CCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGATAAGGATT
GACAG
GTTGAGTGTCCTTTCTCGATTTCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCGATCCTCGGGGGCGATATT
C
AGAGTAGGAAGACCGATAGCGAACAAGTAGCGTGAGCGAAAGTTTGAAAAGTACTTTGGAGAGAGAGTGACATAG
AACCT
GAAGTCGTGACAATGACAATGGAAGTACCTCCATTTCAACGGCCAAAACCAGCCTCATACATCCCGCTGGTATCA
ATCCA
GCGGGCCATGGTGAATCCCTTCTCCGTGTCCTTGATGTTGGCGAAGGAAATGGGGTGCCCCCACCCGTCTTGAAA
CACGG
ACCAAGGAGTCAAACAGACGCGCAATTGCCAGTACTTCTCTCGTACTGAAAGGGAGATGCGAATGCATCTGTGGT
GTGCA
CGCCCAACACCGGCCGGCCCCGGTTGAGTTGGAGTGTGCCTGTTTGGACCCGAAAGGTGGTGAACTATGCCTGAA
CAAGG
TGAAGCCCGGGGAAACCCAGGTGGAGGCCTGTAGCGATGCTGACGTGCAAATCGCTCGCACGATTTGGGTATCGG
AGCGA
AAGACTCATCGAACCACCTAGTAGCTGGTTCACATCGAAGTTTCCCTCAGGATAGCTGGTGCTTGTTTCGGTAAA
GCAAA
TGATTAGAGGCATTGGTGTTCTTAGAATATCGACCTATTCTCAAACTTTAAATGTGCAAAGAAACCGACAAGGGC
GGCAG
ACAAGAATTAAGGCACCAAGTGGGCCTTCTTCGGTAAACGGAACAGGCAATGCGAAATGAACCCCTTAGGGATAT
GTGGG
GTCAAAACTTGTGGGCTCATTCTGATACGTGAAAAGGGTGTTGGTGGATAGGGACAGTTGGACGGTGGCCATGGA
AGTCG
GCATCCGCTAAGGAGTGTGTAACAACTCACCAACCGAATCCACCGG

>gi|148529690|gb|EF375882.1|_Trypanosoma_sp._carpio_18S_ribosomal_RNA_gene,
_part
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTCCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATGAGCTACCGTTTCGGCTTTTGTTGGTTTTAAAAACTCATTGGAGATTATGGAGTTGTGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
GGCGA

AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTT
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATT
C

>gi|160878089|gb|EU185634.1|_Trypanosoma_ophiocephali_18S_small_subunit_rib
osoma
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTCCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATGAGCTACCGTTTCGGCTTTTGTTGGTTTTAAAAACTCATTGGAGATTATGGAGTTGTGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTT
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATT
C

>gi|4468767|emb|AJ009161.1|_Trypanosoma_rotatorium_18S_rRNA_gene,_isolate_B
2-II
GGGCGTCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTCCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATGGGCCACCGTTTCGGCTTTTGTTGGTTTTAAACGTCCATTGGAGATTATGGGGCTGTGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACTAAAGTATGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAT
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTGTTACGGGAATATCCTCAGCA
CGTTT
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGGATTGTCCATAGGATAGCAATCCCTGGGGGCGGTATC
C
AGAGTAGGAAGACCGATAGCGAACAAGTAGCGTGAGCGAAAGTTTGAAAAGCACTTTGGAAAGAGAGTGACATAG
AACCT
GAAGTCGTGACAATGACAATGGAAGTACCTCCATTTCAACGGCCAAAACCAGCCTCACATTTCCGGCGGGTTCAA
ATCCG
TCGGGTCATGGTGAATTTCTTCTCCGTGTCCTCGATGTTGGCGAAGGAAGTGGGGTGCCCCCACCCGTCTTGAAA
CACGG

ACCAAGGAGTCAAACAGACGCGCAATTGCCAGTACTTCTCTCGTACTGAAAGGGAGATGCGATTGCATCTGTGGT
GTGCA
CGCCCAACACCGACCGGCCCCGGTTGAGTTGGAGTGTGCCTGTTTGGACCCGAAAGGTGGTGAACTATGCCTGAA
CAAGG
TGAAGCCCGGGGAAACCCAGGTGGAGGCCTGTAGCGATGCTGACGTGCAAATCGCTCGCATGATTTGGGTATCGG
AGCGA
AAGACTCATCGAACCACCTAGTAGCTGGTTCACGTCGAAGTTTCCCTCAGGATAGCTGGTGCTTGTTTCGGTAAA
GCAAA
TGATTAGAGGCATTGGTGTTCTTAGAGCATCGACCTATTCTCAAACTTTAAATGTGCAAACAAACCTGCAAGGTC
GGTAG
AAAAGAATTAAGGCACCAAGTGGGCCTTCTTCGGTAAGCGGAGCAGGCGATGCGGAATGAACCGCTTAGGGATAT
GTGGC
GTCAAAACTTATGGGCTCATTCTGATACGTGAAAAGGGTGTTGGTGGATAGGGACAGTTGGGCGGTGGCCATGGA
AGTCG
GCACCCGCTAAGGAGTGTGTAACAACTCACCAACCGAATCCACCGG

>gi|2808546|emb|AJ223567.1|_Trypanosoma_mega,_small_subunit_ribosomal_RNA_g
ene
GGGCGTCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTCCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATGGGCCACCGTTTCGGCTTTTGTTGGTTTTAAAAGTCCATTGGAGATTATGGGGCTGTGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTATAACGGGAATATCCTCAGCA
CGTTT
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGGATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATC
C

>gi|4468766|emb|AJ009160.1|_Trypanosoma_rangeli_18S_rRNA_gene,_isolate_RGB_
(Base
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATGGGCCACCGTTTCGGCTTTTGTTGGTTTTAAAAGTCCATTGGAGATTATGGGGCAGTGTGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTAACGGGAATATCCTCAGCA
CGTTA
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATT
C
AGAGTAGGAAGACCGATAGCGAACAAGTAGCGTGAGCGAAAGTTTGAAAAGCACTTTGGAAAGAGAGTGACATAG
AACCT

GAAGTCGTGACAATGACAATGGAAGTACCTCCATTTCAACGGCCAAAACCAGCCTTACATATCTTGTCGGTTTAA
ATCCG
GCAAGTTGTGGTGTCTTTCTTCTCCGTGTCCTCGAATTTGGCGAAGGAAATGGGGTGCCCCCACCCGTCTTGAAA
CACGG
ACCAAGGAGTCAAACAGACGCGCAATTGCCAGTACTTCTCTCGTACTGAAAGGGAGATGCGTATGCATCTGTGGT
GTGCA
CGCCCAACACCGGCCGGCCCCGGTTGAGTTGGAGTGTGCCTGTTTGGACCCGAAAGGTGGTGAACTATGCCTGAA
CAAGG
TGAAGCCCGGGGAAACCCGGGTGGAGGCCTGTAGCGATGCTGACGTGCAAATCGCTCGTACGATTTGGGTATCGG
AGCGA
AAGACTCATCGAACCACCTAGTAGCTGGTTCACATCGAAGTTTCCCTCAGGATAGCTGGTGCTTGTTTCGGTAAA
GCAAA
TGATTAGAGGCATTGGTGTTCTTAGAATATCGACCTATTCTCAAACTTTAAATGTGCAAACAAACCGGCAAGGTC
GGCAG
ACAAGAATTAAGGCACCAAGTGGGCCTTCTTCGGTAAGCGGAGCAGGCGATGCGGAATGAACCGCTTAGGGATAT
GTGGC
GTCAAAACTTATGGGCTCATTCTGATACGTGAAAAGGGTGTTGGTGGATAGGGACAGTTGGACGGTGGCCATGAA
AGTCG
GCATCCGCTAAGGAGTGTGTAACAACTCACCAACCGAATCCACCGG

>gi|320443936|gb|FJ001666.2|_Trypanosoma_dionisii_culture-
collection_TCC/USP:211
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATGGGCCACCGTTTCGGCTTTTGTTGGTTTTAAAAGTCCATTGGAGATTATGGGGCAGTGTGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATT
C

>gi|320443935|gb|FJ001665.2|_Trypanosoma_cruzi_marinkellei_culture-
collection_TC
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATGGGCCACCGTTTCGGCTTTTGTTGGTTTTAAAAGTCCATTGGAGATTATGGGGCAGTGTGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC

GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATT
C

>gi|4468746|emb|AJ009147.1|_Trypanosoma_cruzi_18S_rRNA_gene,_isolate_Sylvio
_X10
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTCTGGGCCACCGTTTCGGCTTTTGTTGGTTTTAAAAGTCCATTGGAGATTATGGGGCAGTGTGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATT
C

>gi|4468773|emb|AJ009163.1|_Trypanosoma_theileri_18S_rRNA_gene,_isolate_TRE
U_124
GGGCGTCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATGGGCCACCGTTTCGGCTTTTGTTGGTTTTAAAAGTCCATTGGAGATTATGGCGTCGTGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGCGCGTCAGAGGTGAAATTCTTAGACCGCGCCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAATTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCCAGGGGCGATATT
C

>gi|301793082|emb|FN666409.1|_Trypanosoma_melophagium_partial_18S_rRNA_gene
,_iso
GGGCGTCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATGGGCCACCGTTTCGGCTTTTGTTGGTTTTAAAAGTCCATTGGAGATTATGGCGTCGTGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGCGCGTCAGAGGTGAAATTCTTAGACCGCGCCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAATTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG

ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCCAGGGGCGATATT
C

>gi|46091659:375-
2593_Trypanosoma_grosi_genes_for_IGS,_18S_rRNA,_5.8S_rRNA,_28S
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATAGGCCACCGTTTCGGCTTTTGTTGGTTTTAACAGTCTATTGGAGATTATGGGGCTGTGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATT
C
AGAGTAGGAAGACCGATAGCGAACAAGTAGCGTGAGCGAAAGTTTGAAAAGCACTTTGGAAAGAGAGTGACATAG
AACCT
GAAGTCGTGACAATGACAATGGAAGTACCTCCATTTCAACGGCCAAAACCAGCCTTACATATCTCTTCGGTGTGT
ACCCG
TTGAGTCATGGTGTATTTCTTCTCCGTGTTCTCGAATTTGGCGAAGGAAATGGGGTGCCCCCACCCGTCTTGAAA
CACGG
ACCAAGGAGTCAAACAGACGCGCAATTGCCAGTACTTCTCTCGTACTGAAAGGGAGATGCATTTGCATCTGTGGT
GTGCA
CGCCCAACACCGGCCGGCCCCGGTTGAGTTGGAGTGTGCCTGTTTGGACCCGAAAGGTGGTGAACTATGCCTGAA
CAAGG
TGAAGCCCGGGGAAACCCAGGTGGAGGCCTGTAGCGATGCTGACGTGCAAATCGCTCGTACGATTTGGGTATCGG
AGCGA
AAGACTCATCGAACCACCTAGTAGCTGGTTCACATCGAAGTTTCCCTCAGGATAGCTGGTGCTTGTTTCGGTAAA
GCAAA
TGATTAGAGGCATTGGTGTTCTTAGAATATCGACCTATTCTCAAACTTTAAATGTGCAAACAAACCGGCAAGGTC
GGAGG
ACAAGAATTAAGGCACCAAGTGGGCCTTCTTCGGTAAGCGGAGCAGGCGATGCGGAATGAACCGCTTAGGGATAT
GTGGC
GTCAAAACTTATGGGCTCATTCTGATGCATGAAAAGGGTGTTGGTGGATAGGGACAGTTGGACGGTGGCCATGGA
AGTCG
GCATCCGCTAAGGAGTGTGTAACAACTCACCAACCGAATCCACCGG

>gi|2808495|emb|AJ223566.1|_Trypanosoma_lewisi,_small_subunit_ribosomal_RNA
_gene
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATAGGCCACCGTTTCGGCTTTTGTTGGTTTTAACAGTCTATTGGAGATTATGGGGCTGTGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA

CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATT
C

>gi|2808547|emb|AJ223568.1|_Trypanosoma_musculi,_small_subunit_ribosomal_RN
A_gen
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATAGGCCACCGTTTCGGCTTTTGTTGGTTTTAACAGTCTATTGGAGATTATGGGGCTGTGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATT
C

>gi|4468759|emb|AJ009158.1|_Trypanosoma_microti_18S_rRNA_gene,_isolate_TRL_
132
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATAGGCCACCGTTTCGGCTTTTGTTGGTTTTAACAGTCTATTGGAGATTATGGGGCTGTGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATT
C

>gi|46091661:28-
2242_Trypanosoma_otospermophili_genes_for_IGS,_18S_rRNA,_5.8S_rR
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATAGGCCACCGTTTCGGCTTTTGTTGGTTTTAACAGTCTATTGGAGATTATGGGGCTGTGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA

TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATT
C
AGAGTAGGAAGACCGATAGCGAACAAGTAGCGTGAGCGAAAGTTTGAAAAGCACTTTGGAAAGAGAGTGACATAG
AACCT
GAAGTCGTGACAATGACAATGGAAGTACCTCCATTTCAACGGCCAAAACCAGCCTTACATATCTCTTCGGTGCGT
ACCCG
TTGAGTCATGGTGTATTTCTTCTCCGTGTTCTCGAATTTGGCGAAGGAAATGGGGTGCCCCCACCCGTCTTGAAA
CACGG
ACCAAGGAGTCAAACAGACGCGCACTTGCCAGTACTTCTCTCGTACTGAAAGGGAGATGCGTTTGCATCTGTGGT
GTGCA
CGCCCAACACCGGCCGGCCCCGGTTGAGTTGGAGTGTGCCTGTTTGGACCCGAAAGGTGGTGAACTATGCCTGAA
CAAGG
TGAAGCCCGGGGAAACCCAGGTGGAGGCCTGTAGCGATGCTGACGTGCAAATCGCTCGCACGATTTGGGTATCGG
AGCGA
AAGACTCATCGAACCACCTAGTAGCTGGTTCACATCGAAGTTTCCCTCAGGATAGCTGGTGCTTGTTTCGGTAAA
GCAAA
TGATTAGAGGCATTGGTGTTCTTAGAATATCGACCTATTCTCAAACTTTAAATGTGCAAACAAACCGGCAAGGTC
GGAGG
ACAAGAATTAAGGCACCAAGTGGGCCTTCTTCGGTAAGCGGAGCAGGCGATGCGGAATGAACCGCTTAGGGATAT
GTGGC
GTCAAAACTTATGGGCTCATTCTGATGCATGAAAAGGGTGTTGGTGGATAGGGACAGTTGGACGGTGGCCATGGA
AGTCG
GCATCCGCTAAGGAGTGTGTAACAACTCACCAACCGAATCCACCGG

>gi|46091662:28-
2245_Trypanosoma_kuseli_genes_for_IGS,_18S_rRNA,_5.8S_rRNA,_28S
GGGCGCCCGTGATTTTTACTGTGACTAAAAAAGTGTGACCAAAGCAGTCTTCGACTTGAATTAGAAAGCATGGGA
TAACA
AAGGAGCAGCCTATAGGCCACCGTTTCGGCTTTTGTTGGTTTTAACAGTCTATTGGAGATTATGGGGCTGTGCGA
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACGATGACACCCATGAATTGGGGAGTTTTAATTTACGTGCATATTCTTTTACGGGAATATCCTCAGCA
CGTTA
TCTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGATTCAGAATTGCCCATAGGATAGCAATCCCTGGGGGCGGTATT
C
AGAGTAGGAAGACCGATAGCGAACAAGTAGCGTGAGCGAAAGTTTGAAAAGCACTTTGGAAAGAGAGTGACATAG
AACCT
GAAGTCGTGACAATGACAATGGAAGTACCTCCATTTCAACGGCCAAAACCAGCCTTACATATCTCTTCGGTGTGT
ACCCG
TTGAGTCATGGTGTATTTCTTCTCCGTGTTCTCGAATTTGGCGAAGGAAATGGGGTGCCCCCACCCGTCTTGAAA
CACGG
ACCAAGGAGTCAAACAGACGCGCAATTGCCAGTACTTCTCTCGTACTGAAAGGGAGATGCATTTGCATCTGTGGT
GTGCA
CGCCCAACACCGGCCGGCCCCGGTTGAGTTGGAGTGTGCCTGTTTGGACCCGAAAGGTGGTGAACTATGCCTGAA
CAAGG

TGAAGCCCGGGGAAACCCAGGTGGAGGCCTGTAGCGATGCTGACGTGCAAATCGCTCGTACGATTTGGGTATCGG
AGCGA
AAGACTCATCGAACCACCTAGTAGCTGGTTCACATCGAAGTTTCCCTCAGGATAGCTGGTGCTTGTTTCGGTAAA
GCAAA
TGATTAGAGGCATTGGTGTTCTTAGAATATCGACCTATTCTCAAACTTTAAATGTGCAAGCAAACCGGCAAGGTC
GGAGG
ACAAGAATTAAGGCACCAAGTGGGCCTTCTTCGGTAAGCGGAGCAGGCGATGCGGAATGAACCGCTTAGGGATAT
GTGGC
GTCAAAACTTATGGGCTCATTCTGATGCATGAAAAGGGTGTTGGTGGATAGGGACAGTTGGACGGTGGCCATGGA
AGTCG
GCATCCGCTAAGGAGTGTGTAACAACTCACCAACCGAATCCACCGG
>gi|4468741|emb|AJ009142.1|_Trypanosoma_brucei_rhodesiense_18S_rRNA_gene,_i
solat
GGGCGTCCGTGAATTTTACTGTGACCAAAAAAGTGCGACCAAAGCAGTCGCCGACTTGAATTACAAAGCATGGGA
TAACG
AAGCATCAGCCCTGGGGCCACCGTTTCGGCTTTTGTTGGTTTTAGAAGTCCTTGGGAGATTATGGGGCCGCGTGC
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCAAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACCATGACACCCATGAATTGGGGAACATCAATTTACGTGCCTATTCTATCACGGGAATATCCTCAGCA
CGTTT
CTTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATGGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGTGCCGGGATTGTCCACAGGACAGCAGTCCCTGTGGTCGTCATC
C

>gi|4468740|emb|AJ009141.1|_Trypanosoma_brucei_gambiense_18S_rRNA_gene,_iso
late
GGGCGTCCGTGAATTTTACTGTGACCAAAAAAGTGCGACCAAAGCAGTCGCCGACTTGAATTACAAAGCATGGGA
TAACG
AAGCATCAGCCCTGGGGCCACCGTTTCGGCTTTTGTTGGTTTTAGAAGTCCTTGGGAGATTATGGGGCCGCGTGC
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCAAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACCATGACACCCATGAATTGGGGAACATCAATTTACGTGCCTATTCTATCACGGGAATATCCTCAGCA
CGTTT
CTTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATGGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGTGCCGGGATTGTCCACAGGACAGCAGTCCCTGTGGTCGTCATC
C

>gi|4468753|emb|AJ009154.1|_Trypanosoma_evansi_18S_rRNA_gene,_isolate_E_110
GGGCGTCCGTGAATTTTACTGTGACCAAAAAAGTGCGACCAAAGCAGTCGCCGACTTGAATTACAAAGCATGGGA
TAACG
AAGCATCAGCCCTGGGGCCACCGTTTCGGCTTTTGTTGGTTTTAGAAGTCCTTGGGAGATTATGGGGCCGCGTGC
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA

AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCAAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACCATGACACCCATGAATTGGGGAACATCAATTTACGTGCCTATTCTATCACGGGAATATCCTCAGCA
CGTTT
CTTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATGGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGTGCCGGGATTGTCCACAGGACAGCAGTCCCTGTGGTCGTCATC
C

>gi|4468768|emb|AJ009162.1|_Trypanosoma_simiae_18S_rRNA_gene,_isolate_KEN_2
GGGCGTCCGTGATCTTTACTGTGACTAAAAAAGAGCGACCAAAGCAGTCCACGACTTGAATTACAAAGCATGGGA
TAACA
AAGCATCAGCCCCAGGGCCACCGTTTCGGCTTTTGTTGGTTTTAGAAGCCCAGGGGAGATTATGGCGCCGCGTGC
CGGAA
TGAATGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGTATCGAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACCATGACACCCATGAATTGGGGACCATTAACTTACGTGCCTATTCTGCCACGGGAATATCCTCAGCA
CGTTG
TTTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATGGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGGCCCGTGATCGTCCACAGGACAGCATTCCGTCGGAGCACTATC
A
AGAGTAGGAAGACCGATAGCGAACAAGTAGCGTGAGCGAAAGTTTGAAAAGCACTTTGGAAAGAGAGTGACATAG
AACCT
GAAATCGTGACAACGGCAATGGAAGTACCTCCATTTCGACGGCCAGACCCAGCCGCGCGGGCGCCGGGGGTTCGA
ATCCC
GGCGGGCGTGGCGTGTGCCCTGTCCGTGGTCTCGACGCTGGCGAAGGGAATGGGGTGCCCCCACCCGTCTTGAAA
CACGG
ACCAAGGAGTCAAACAGACGCGCAAGTATGAGAACCTCTCTCGTACTGAAAGGGAGATGCGAAGGCATCCGTGGT
GTGCG
CGCTCAACGCCGGCCGGCCCCGGCTGCGCGCGAGTGTGCCTGTTTGGACCCGAAAGGTGGTGAACTATGCCTGAA
CAAGG
TGAAGCCCGGGGAAACCCCGGTGGAGGCCTGTAGCGATGCTGACGTGCAAATCGCTCGTACGATTTGGGTATCGG
AGCGA
AAGACTCATCGAACCACCTAGTAGCTGGTTCACGTCGAAGTTTCCCTCAGGATAGCTGGTGCTTGTTCCGGTAAA
GCAAA
TGATTAGAGGCATTGGTGTTCTTGTAATATCGACCTATTCTCAAACTTTAAATGTGCAAGAAGCTCGGCAAGGCC
GGCGG
GCAAGAATTAAGGCACCAAGTGGGCCTTCTTCGGTAAGCGGAGCAGGCGATGCGGAATGAACCGCTTAGGGATAT
GTGGC
GTCGAAACTCGTGGGCTCATTCGGATACGTGAAAAGGGTGTTGGTGGATAGGGACAGTTGGACGGTGGCCATGGA
AGTCG
GCATCCGCTAAGGAGTGTGTAACAACTCACCAACCGAATCCACCGG

>gi|4468743|emb|AJ009144.1|_Trypanosoma_congolense_(kilifi)_18S_rRNA_gene,_
isola
GGGCGTCCGTGATTTTTACTGTGACTAAAAAAGTGCGACCAAAGCAGTCGCCGACTTGAATTACAAAGCATGGGA
TAACA
AAGCACCAGCCCTAGGGCCACCGTTTCGGCTTTTGTTGGTTTTAGAAGTCCTTGGGAGATTATGGTGTCGCGTGC
CGGAA

TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCAAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACCATGACACCCATGAATTGGGGACCACCAACTTACGTGTCTATTCTGCTACGGGAATATCCTCAGCA
CGTTG
TTTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGGCCCGTGATCGTCCACAGGACAGCCTACCGTGGCGGCGCTATC
A
AGAGTAGGAAGACCGATAGCGAACAAGTAGCGTGAGCGAAAGTTTGAAAAGCACTTTGGAAAGAGAGTGACATAG
AACCT
GAAATCGTGACAACGACAATGGAAGTACCTCCATTTCGACGGCCAGACCCAGCCGCGCGTGCCTCGGGGGTCAAA
ATCCC
GTGGGGCGTGGCGTGTGCTCTTTCCGTGGTCTCGACGCTGGCGAAGGGAATGGGGTGCCCCCACCCGTCTTGAAA
CACGG
ACCAAGGAGTCAAACAGACGCGCAATTGGGAGAACCTCTCCCGTACTGAAAGGGAGATGCGAATGCATCTGTGGT
GTGCG
CGCCCAATGCCGGCCGGCCCCGGCTGCGCGTGAGTGTGCCTGTTTGGACCCGAAAGGTGGTGAACTATGCCTGAA
CAAGG
TGAAGCCCGGGGAAACCCCGGTGGAGGCCTGTAGCGATGCTGACGTGCAAATCGCTCGTACGATTTGGGTATCGG
AGCGA
AAGACTCATCGAACCACCTAGTAGCTGGTTCACGTCGAAGTTTCCCTCAGGATAGCTGGTGCTCGTCTCGGTAAA
GCAAA
TGATTAGAGGCATTGGTGTTCTTGTAATATCGACCTATTCTCAAACTTTAAATGTGCAAGAAGCGCGGCAAGGCC
GGCAG
ACAAGAATTGAGGCACCAAGTGGGCCTTCTTCGGTAAGCGGAGCAGGCGATGCGGAATGAACCGCTTAGGGATAT
GTGGC
GTCGAAACTCGTGGGCTCATTCTACTTCGTGAAAAGGGTGTTGGTGGATAGGGACAGTTGGACGGTGGCCATGGA
AGTCG
GCATCCGCTAAGGAGTGTGTAACAACTCACCAACCGAATCCACCGG

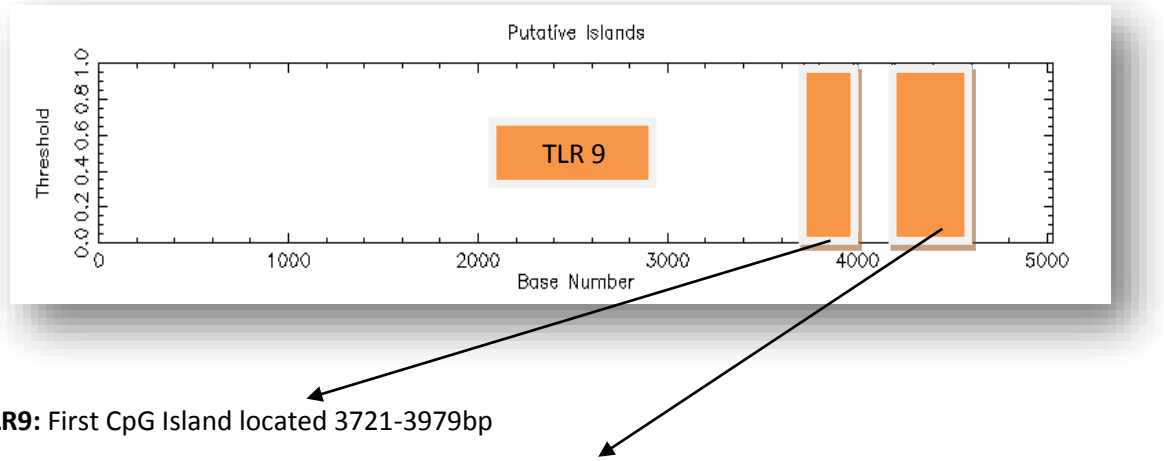
>gi|4468745|emb|AJ009146.1|_Trypanosoma_congolense_(savannah)_18S_rRNA_gene
,_iso
GGGCGTCCGTGATTTTTACTGTGACTAAAAAAGTGCGACCAAAGCAGTCGCCGACTTGAATTACAAAGCATGGGA
TAACA
AAGCACCAGCCCTAGGGCCACCGTTTCGGCTTTTGTTGGTTTTAGAAGTCCTTGGGAGATTATGGCCTCGCGTGC
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCAAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACCATGACACCCATGAATTGGGGACCACCAACTTACGTGTCTATTCTGCTACGGGAATATCCTCAGCA
CGTTG
TTTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGGCCCGTGATTGTCCACAGGACAGCCTACCGTGGCGGCGCTATC
A
AGAGTAGGAAGACCGATAGCGAACAAGTAGCGTGAGCGAAAGTTTGAAAAGCACTTTGGAAAGAGAGTGACATAG
AACCT
GAAATCGTGACAACGACAATGGAAGTACCTCCATTTCGACGGCCAGACCCAGCCGCGCGTGCCTTGGGGGTCAAA
ATCCC

TTGGGGCGTGGCGTGTGCTCTTTCCGTGCTCTCGACGCTGGCGAAGGGAATGGGGTGCCCCCACCCGTCTTGAAA
CACGG
ACCAAGGAGTCAAACAGACGCGCAATTGTGAGAACCTCTCTCGTACTGAAAGGGAGATGCGAATGCATCCGTGGC
GTGCG
CGCCCAACGCCGGCCGGCCCCGGCTGCGCGTGAGTGTGCCTGTTTGGACCCGAAAGGTGGTGAACTATGCCTGAA
CAAGG
TGAAGCCCGGGGAAACCCCGGTGGAGGCCTGTAGCGATGCTGACGTGCAAATCGCTCGTACGATTTGGGTATCGG
AGCGA
AAGACTCATCGAACCACCTAGTAGCTGGTTCACGTCGAAGTTTCCCTCAGGATAGCTGGTGCTCGTGTCGGTAAA
GCAAA
TGATTAGAGGCATTGGTGTTCTTGTAATATCGACCTATTCTCAAACTTTAAATGTGCAAGAAGCTCGGCAAGGCC
GGCAG
ACAAGAATTGAGGCACCAAGTGGGCCTTCTTCGGTAAGCGGAGCAGGCGATGCGGAATGAACCGCTTAGGGATAT
GTGGC
GTCGAAACTCGTGGGCTCGTTCTGATACGTGAAAAGGGTGTTGGTGGATAGGGACAGTTGGACGGTGGCCATGGA
AGTCG
GCATCCGCTAAGGAGTGTGTAACAACTCACCAACCGAATCCACCGG

>gi|4468744|emb|AJ009145.1|_Trypanosoma_congolense_(riverine_forest)_18S_rR
NA_ge
GGGCGTCCGTGATTTTTACTGTGACTAAAAAAGTGCGACCAAAGCAGTCGCCGACTTGAATTACAAAGCATGGGA
TAACA
AAGCACCAGCCCTGGGGCCACCGTTTCGGCTTTTGTTGGTTTTAGAAGTCCTTGGGAGATTATGGCGCCGCGTGC
CGGAA
TGAAGGAGGGTAGTCGGGGGAGAACGTACTGGTGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTAC
AGCGA
AGGCATTCTTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGGGATCAAAGATGATTAGAGACCATTGTAG
TCCAC
ACTGCAAACCATGACACCCATGAATTGGGGACCACCAACTTACGTGTCTATTCTGCTACGGGAATATCCTCAGCA
CGTTG
TTTACGCGAAAGCTTAGGTTACAGTCTCAGGGGGGAGTACGTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAA
TGGCA
CCACAAGACGTGGAGCGTGCGGTTTAATTTGACTCAACACGGGGAACTTTACCAGATCCGGACAGGGTGAGGATT
GACAG
ATTGAGTGTTCTTTCTCGATCCCCTGAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTG
ATTCC
GTCAACGGACGAGATCCAAGCTGCCCAGTAGGGCCCGTGATTGTCCACAGGACAGCCTACCGTGGCGGCGCTATC
A
AGAGTAGGAAGACCGATAGCGAACAAGTAGCGTGAGCGAAAGTTTGAAAAGCACTTTGGAAAGAGAGTGACATAG
AACCT
GAAATCGTGACAACGGCAATGGAAGTACCTCCATTTCGACGGCCAGACCCAGCCGCGCGTGCCTCGGGGGTCAAA
ATCCC
TTGGGGCGTGGCGTGTGCTCTTTCCGTGCTCTCGACGCTGGCGAAGGGAATGGGGTGCCCCCACCCGTCTTGAAA
CACGG
ACCAAGGAGTCAAACAGACGCGCAATTGTGAGAACCTCTCTCGTACTGAAAGGGAGATGCGAATGCATCTGTGGC
GTGCG
CGCCCAACGCCGGCCGGCCCCGGCTGCGCGTGAGTGTGCCTGTTTGGACCCGAAAGGTGGTGAACTATGCCTGAA
CAAGG
TGAAGCCCGGGGAAACCCCGGTGGAGGCCTGTAGCGATGCTGACGTGCAAATCGCTCGTACGATTTGGGTATCGG
AGCGA
AAGACTCATCGAACCACCTAGTAGCTGGTTCACGTCGAAGTTTCCCTCAGGATAGCTGGTGCTCGTGTCGGTAAA
GCAAA
TGATTAGAGGCATTGGTGTTCTTGTAATATCGACCTATTCTCAAACTTTAAATGTGCAAGAAGCGCGGCAAGGCC
GGCAG
ACAAGAATTGAGGCACCAAGTGGGCCTTCTTCGGTAAGCGGAGCAGGCGATGCGGAATGAACCGCTTAGGGATAT
GTGGC
GTCGAAACTCGTGGGCTCATTCTGATACGTGAAAAGGGTGTTGGTGGATAGGGACAGTTGGACGGTGGCCATGGA
AGTCG
GCATCCGCTAAGGAGTGTGTAACAACTCACCAACCGAATCCACCGG

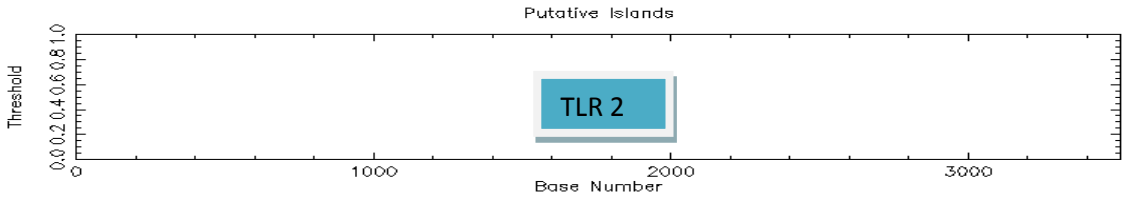
Appendix F: CpG Graph plots for TLR9, TLR2, TLR4 and TLR6

The four plots shown below shows predicted CpG Islands using CpgEMBOSS plot. TLR9 has 2 Islands while TLR2, 4 and 6 have no Islands (See Chapter 5).

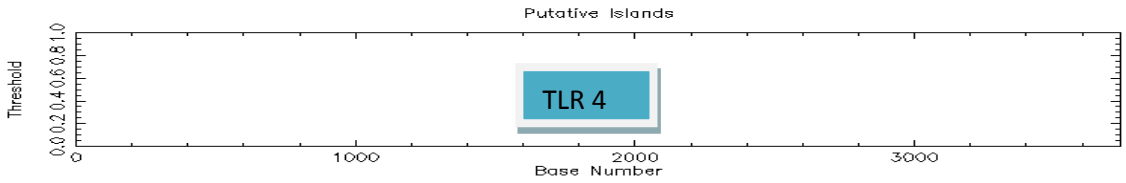


TLR9: First CpG Island located 3721-3979bp

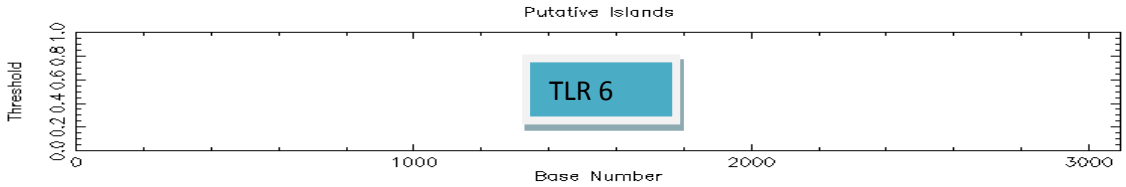
Second CpG Island located 4182-4602bp



TLR2: No CpG Islands



TLR4: No CpG Islands



TLR6: No CpG Islands

Appendix G: Multiple sequence alignment of bovine TLR9 genes

Alignment generated using Clusta Omega Software (Chapter 5)

NM_183081.1: Bos taurus toll-like receptor 9 (TLR9), Mrna
FJ495080.2: Bos indicus breed Sahiwal toll-like receptor 9 (TLR9) gene, complete cds
EF076726.1: Bos indicus breed Brahman toll-like receptor 9 gene, complete cds
EF076731.1: Bos taurus breed Piedmontese toll-like receptor 9 gene, complete cds
EF076727.1: Bos taurus breed Charolais toll-like receptor 9 gene, complete cds
EF076725.1: Bos taurus breed Braford toll-like receptor 9 gene, complete cds
EF076724.1: Bos taurus breed Angus toll-like receptor 9 gene, complete cds
EF076729.1: Bos taurus breed Limousin toll-like receptor 9 gene, complete cds
EF076728.1: Bos taurus breed Holstein toll-like receptor 9 gene, complete cds
EF076723.1: Bos taurus breed Romagnola toll-like receptor 9 gene, complete cds
AJ509825.1: Bos taurus partial mRNA for toll-like receptor 9 (tlr9 gene)
AJ509824.2: Bos taurus partial tlr9 gene for Toll-like receptor 9
AY859726.1: Bos taurus toll-like receptor 9 mRNA, complete cds

```
gi|407316560|gb|FJ495080.2|       ---------------AGGGCCTGCTCCTGGAGGAAGCTGCCGCTGTGGCAGGGGCAGATG     45
gi|126033143|gb|EF076726.1|       CTGGGGATGCAGAGGAGGGCCTGCTCCTGGAGGAAGCTGCCGCTGTGGCAGGGGCAGATG     60
gi|126033153|gb|EF076731.1|       CTGGGGATGCAGAGGAGGGCCTGCTCCTGGAGGAAGCTGCCGCTGTGGCAGGGGCAGATG     60
gi|126033145|gb|EF076727.1|       CTGGGGATGCAGAGGAGGGCCTGCTCCTGGAGGAAGCTGCCGCTGTGGCAGGGGCAGATG     60
gi|126033141|gb|EF076725.1|       CTGGGGATGCAGAGGAGGGCCTGCTCCTGGAGGAAGCTGCCGCTGTGGCAGGGGCAGATG     60
gi|126033139|gb|EF076724.1|       CTGGGGATGCAGAGGAGGGCCTGCTCCTGGAGGAAGCTGCCGCTGTGGCAGGGGCAGATG     60
gi|126033149|gb|EF076729.1|       CTGGGGATGCAGAGGAGGGCCTGCTCCTGGAGGAAGCTGCCGCTGTGGCAGGGGCAGATG     60
gi|126033147|gb|EF076728.1|       CTGGGGATGCAGAGGAGGGCCTGCTCCTGGAGGAAGCTGCCGCTGTGGCAGGGGCAGATG     60
gi|126033137|gb|EF076723.1|       CTGGGGATGCAGAGGAGGGCCTGCTCCTGGAGGAAGCTGCCGCTGTGGCAGGGGCAGATG     60
gi|57471257|gb|AY859726.1|        ------------------------------------------------------------     0
gi|27848078|emb|AJ509824.2|       ------------------------------------------------------------     0
gi|34147166|ref|NM_183081.1|      ------------------------------------------------------------     0
gi|23821159|emb|AJ509825.1|       ------------------------------------------------------------     0

gi|407316560|gb|FJ495080.2|       ACCCTGGGCAAGGCCAGTGCTGCCCCCTGCCCCATACGCCAGGCCTGGCATGACACACAG     105
gi|126033143|gb|EF076726.1|       ACCCTGGGCAAGGCCAGTGCTGCCCCCTGCCCCATACGCCAGGCCTGGCATGACACACAG     120
gi|126033153|gb|EF076731.1|       ACCCTGGGCAAGGCCAGTGCTGCCCCCTGCCCCATACGCCAGGCCTGGCATGACACACAG     120
gi|126033145|gb|EF076727.1|       ACCCTGGGCAAGGCCAGTGCTGCCCCCTGCCCCATACGCCAGGCCTGGCATGACACACAG     120
```

```
gi|126033141|gb|EF076725.1|       ACCCTGGGCAAGGCCAGTGCTGCCCCCTGCCCCATACGCCAGGCCTGGCATGACACACAG     120
gi|126033139|gb|EF076724.1|       ACCCTGGGCAAGGCCAGTGCTGCCCCCTGCCCCATACGCCAGGCCTGGCATGACACACAG     120
gi|126033149|gb|EF076729.1|       ACCCTGGGCAAGGCCAGTGCTGCCCCCTGCCCCATACGCCAGGCCTGGCATGACACACAG     120
gi|126033147|gb|EF076728.1|       ACCCTGGGCAAGGCCAGTGCTGCCCCCTGCCCCATACGCCAGGCCTGGCATGACACACAG     120
gi|126033137|gb|EF076723.1|       ACCCTGGGCAAGGCCAGTGCTGCCCCCTGCCCCATACGCCAGGCCTGGCATGACACACAG     120
gi|57471257|gb|AY859726.1|        ------------------------------------------------------------     0
gi|27848078|emb|AJ509824.2|       ------------------------------------------------------------     0
gi|34147166|ref|NM_183081.1|      ------------------------------------------------------------     0
gi|23821159|emb|AJ509825.1|       ------------------------------------------------------------     0

gi|407316560|gb|FJ495080.2|       AGCTCTGGGACTGGTACCCCAGGGCAGGAAAGGACAGTGAAGGAGCTGCCAGCCTGAGAC     165
gi|126033143|gb|EF076726.1|       AGCTCTGGGACTGGTACCCCAGGGCAGGAAAGGACAGTGAAGGAGCTGCCAGCCTGAGAC     180
gi|126033153|gb|EF076731.1|       AGCTCTGGGACTGGTACCCCAGGGCAGGGAAGGACAGTGAAGGAGCTGCCAGCCTGAGAC     180
gi|126033145|gb|EF076727.1|       AGCTCTGGGACTGGTACCCCAGGGCAGGGAAGGACAGTGAAGGAGCTGCCAGCCTGAGAC     180
gi|126033141|gb|EF076725.1|       AGCTCTGGGACTGGTACCCCAGGGCAGGGAAGGACAGTGAAGGAGCTGCCAGCCTGAGAC     180
gi|126033139|gb|EF076724.1|       AGCTCTGGGACTGGTACCCCAGGGCAGGGAAGGACAGTGAAGGAGCTGCCAGCCTGAGAC     180
gi|126033149|gb|EF076729.1|       AGCTCTGGGACTGGTACCCCAGGGCAGGAAAGGACAGTGAAGGAGCTGCCAGCCTGAGAC     180
gi|126033147|gb|EF076728.1|       AGCTCTGGGACTGGTACCCCAGGGCAGGAAAGGACAGTGAAGGAGCTGCCAGCCTGAGAC     180
gi|126033137|gb|EF076723.1|       AGCTCTGGGACTGGTACCCCAGGGCAGGAAAGGACAGTGAAGGAGCTGCCAGCCTGAGAC     180
gi|57471257|gb|AY859726.1|        ------------------------------------------------------------     0
gi|27848078|emb|AJ509824.2|       ------------------------------------------------------------     0
gi|34147166|ref|NM_183081.1|      ------------------------------------------------------------     0
gi|23821159|emb|AJ509825.1|       ------------------------------------------------------------     0

gi|407316560|gb|FJ495080.2|       TGCTGCCTGGGGGGGCAGGGGCTGAGAAGGCTGAGGCATACCCCACCAGGTCTGACACTG     225
gi|126033143|gb|EF076726.1|       TGCTGCCTGGGGGGGCAGGGGCTGAGAAGGCTGAGGCATACCCCACCAGGTCTGACACTG     240
gi|126033153|gb|EF076731.1|       TGCTGCCTGGGGGGGCAGGGGCTGAGAAGGCTGAGGCATACCCCACCAGGTCTGACACTG     240
gi|126033145|gb|EF076727.1|       TGCTGCCTGGGGGGGCAGGGGCTGAGAAGGCTGAGGCATACCCCACCAGGTCTGACACTG     240
gi|126033141|gb|EF076725.1|       TGCTGCCTGGGGGGGCAGGGGCTGAGAAGGCTGAGGCATACCCCACCAGGTCTGACACTG     240
gi|126033139|gb|EF076724.1|       TGCTGCCTGGGGGGGCAGGGGCTGAGAAGGCTGAGGCATACCCCACCAGGTCTGACACTG     240
gi|126033149|gb|EF076729.1|       TGCTGCCTGGGGGGGCAGGGGCTGAGAAGGCTGAGGCATACCCCACCAGGTCTGACACTG     240
gi|126033147|gb|EF076728.1|       TGCTGCCTGGGGGGGCAGGGGCTGAGAAGGCTGAGGCATACCCCACCAGGTCTGACACTG     240
gi|126033137|gb|EF076723.1|       TGCTGCCTGGGGGGGCAGGGGCTGAGAAGGCTGAGGCATACCCCACCAGGTCTGACACTG     240
gi|57471257|gb|AY859726.1|        ------------------------------------------------------------     0
gi|27848078|emb|AJ509824.2|       ------------------------------------------------------------     0
gi|34147166|ref|NM_183081.1|      ------------------------------------------------------------     0
gi|23821159|emb|AJ509825.1|       ------------------------------------------------------------     0
```

```
gi|407316560|gb|FJ495080.2|       CTCCTAATTCCAATCTTTTGCCAACCTTAGACTGAGCTTTGGGCCTTCACCTCTCCCCAG     285
gi|126033143|gb|EF076726.1|       CTCCTAATTCCAATCTTTTGCCAACCTTAGACTGAGCTTTGGGCCTTCACCTCTCCCCAG     300
gi|126033153|gb|EF076731.1|       CTCCTAATTCCAATCTTTTGCCAACCTTAGACTGAGCTTTGGGCCTTCACCTCTCCCCAG     300
gi|126033145|gb|EF076727.1|       CTCCTAATTCCAATCTTTTGCCAACCTTAGACTGAGCTTTGGGCCTTCACCTCTCCCCAG     300
gi|126033141|gb|EF076725.1|       CTCCTAATTCCAATCTTTTGCCAACCTTAGACTGAGCTTTGGGCCTTCACCTCTCCCCAG     300
gi|126033139|gb|EF076724.1|       CTCCTAATTCCAATCTTTTGCCAACCTTAGACTGAGCTTTGGGCCTTCACCTCTCCCCAG     300
gi|126033149|gb|EF076729.1|       CTCCTAATTCCAATCTTTTGCCAACCTTAGACTGAGCTTTGGGCCTTCACCTCTCCCCAG     300
gi|126033147|gb|EF076728.1|       CTCCTAATTCCAATCTTTTGCCAACCTTAGACTGAGCTTTGGGCCTTCACCTCTCCCCAG     300
gi|126033137|gb|EF076723.1|       CTCCTAATTCCAATCTTTTGCCAACCTTAGACTGAGCTTTGGGCCTTCACCTCTCCCCAG     300
gi|57471257|gb|AY859726.1|        ------------------------------------------------------------     0
gi|27848078|emb|AJ509824.2|       ---------------------------------GAGATTTGGGCCTTCACCTCTCCCCAG     27
gi|34147166|ref|NM_183081.1|      ------------------------------------------------------------     0
gi|23821159|emb|AJ509825.1|       ------------------------------------------------------------     0

gi|407316560|gb|FJ495080.2|       ACTTCCCAGAGGACAGGTGTGAGGCCCACGTGGGATGTGACAGATGGTTCAGACAGACTG     345
gi|126033143|gb|EF076726.1|       ACTTCCCAGAGGACAGGTGTGAGGCCCACGTGGGATGTGACAGATGGTTCAGACAGACTG     360
gi|126033153|gb|EF076731.1|       ACTTCCCAGAGGACAGGTGTGAGGCCCACGTGGGATGTGACAGATGGTTCAGACAGACTG     360
gi|126033145|gb|EF076727.1|       ACTTCCCAGAGGACAGGTGTGAGGCCCACGTGGGATGTGACAGATGGTTCAGACAGACTG     360
gi|126033141|gb|EF076725.1|       ACTTCCCAGAGGACAGGTGTGAGGCCCACGTGGGATGTGACAGATGGTTCAGACAGACTG     360
gi|126033139|gb|EF076724.1|       ACTTCCCAGAGGACAGGTGTGAGGCCCACGTGGGATGTGACAGATGGTTCAGACAGACTG     360
gi|126033149|gb|EF076729.1|       ACTTCCCAGAGGACAGGTGTGAGGCCCACGTGGGATGTGACAGATGGTTCAGACAGACTG     360
gi|126033147|gb|EF076728.1|       ACTTCCCAGAGGACAGGTGTGAGGCCCACGTGGGATGTGACAGATGGTTCAGACAGACTG     360
gi|126033137|gb|EF076723.1|       ACTTCCCAGAGGACAGGTGTGAGGCCCACGTGGGATGTGACAGATGGTTCAGACAGACTG     360
gi|57471257|gb|AY859726.1|        ------------------------------------------------------------     0
gi|27848078|emb|AJ509824.2|       ACTTCCCAGAGGACAGGTGTGAGGCCCACGTGGGATGTGACAGATGGTTCAGACAGACTG     87
gi|34147166|ref|NM_183081.1|      ------------------------------------------------------------     0
gi|23821159|emb|AJ509825.1|       ------------------------------------------------------------     0

gi|407316560|gb|FJ495080.2|       GGGTGGGGGGCGGGGAGTAGAAAGAGGAAGGGGTGGCGGAGCGATCTGCCACCTGACTAT     405
gi|126033143|gb|EF076726.1|       GGGTGGGGGGCGGGGAGTAGAAAGAGGAAGGGGTGGCGGAGCGATCTGCCACCTGACTAT     420
gi|126033153|gb|EF076731.1|       GGGTGGGGGGCGGGGAGTAGAAAGAGGAAGGGGTGGCGGAGCGATCTGCCACCTGACTAT     420
gi|126033145|gb|EF076727.1|       GGGTGGGGGGCGGGGAGTAGAAAGAGGAAGGGGTGGCGGAGCGATCTGCCACCTGACTAT     420
gi|126033141|gb|EF076725.1|       GGGTGGGGGGCGGGGAGTAGAAAGAGGAAGGGGTGGCGGAGCGATCTGCCACCTGACTAT     420
gi|126033139|gb|EF076724.1|       GGGTGGGGGGCGGGGAGTAGAAAGAGGAAGGGGTGGCGGAGCGATCTGCCACCTGACTAT     420
gi|126033149|gb|EF076729.1|       GGGTGGGGGGCGGGGAGTAGAAAGAGGAAGGGGTGGCGGAGCGATCTGCCACCTGACTAT     420
```

```
gi|126033147|gb|EF076728.1|       GGGTGGGGGGCGGGGAGTAGAAAGAGGAAGGGGTGGCGGAGCGATCTGCCACCTGACTAT     420
gi|126033137|gb|EF076723.1|       GGGTGGGGGGCGGGGAGTAGAAAGAGGAAGGGGTGGCGGAGCGATCTGCCACCTGACTAT     420
gi|57471257|gb|AY859726.1|        ------------------------------------------------------------     0
gi|27848078|emb|AJ509824.2|       GGGTGGGGGGCGGGGAGTAGAAAGAGGAAGGGGTGGCGGAGCGATCTGCCACCTGACTAT     147
gi|34147166|ref|NM_183081.1|      ------------------------------------------------------------     0
gi|23821159|emb|AJ509825.1|       ------------------------------------------------------------     0


gi|407316560|gb|FJ495080.2|       GCAAATGGCCTCTCTGACTCATTATGTCCCCTACACCCCGCCCCTGCCATGAGCATAGGG     465
gi|126033143|gb|EF076726.1|       GCAAATGGCCTCTCTGACTCATTATGTCCCCTACACCCCGCCCCTGCCATGAGCATAGGG     480
gi|126033153|gb|EF076731.1|       GCAAATGGCCTCTCTGACTCATTATGTCCCCTACACCCCGCCCCTGCCATGAGCATAGGG     480
gi|126033145|gb|EF076727.1|       GCAAATGGCCTCTCTGACTCATTATGTCCCCTACACCCCGCCCCTGCCATGAGCATAGGG     480
gi|126033141|gb|EF076725.1|       GCAAATGGCCTCTCTGACTCATTATGTCCCCTACACCCCGCCCCTGCCATGAGCATAGGG     480
gi|126033139|gb|EF076724.1|       GCAAATGGCCTCTCTGACTCATTATGTCCCCTACACCCCGCCCCTGCCATGAGCATAGGG     480
gi|126033149|gb|EF076729.1|       GCAAATGGCCTCTCTGACTCATTATGTCCCCTACACCCCGCCCCTGCCATGAGCATAGGG     480
gi|126033147|gb|EF076728.1|       GCAAATGGCCTCTCTGACTCATTATGTCCCCTACACCCCGCCCCTGCCATGAGCATAGGG     480
gi|126033137|gb|EF076723.1|       GCAAATGGCCTCTCTGACTCATTATGTCCCCTACACCCCGCCCCTGCCATGAGCATAGGG     480
gi|57471257|gb|AY859726.1|        ------------------------------------------------------------     0
gi|27848078|emb|AJ509824.2|       GCAAATGGCCTCTCTGACTCATTATGTCCCCTACACCCCGCCCCTGCCATGAGCATAGGG     207
gi|34147166|ref|NM_183081.1|      ------------------------------------------------------------     0
gi|23821159|emb|AJ509825.1|       ------------------------------------------------------------     0


gi|407316560|gb|FJ495080.2|       AGGAAGGGGGCTGCTAGGTTATTACAAAATTCTCACTTCCTCTGATCTCTGAGCCGGCGG     525
gi|126033143|gb|EF076726.1|       AGGAAGGGGGCTGCTAGGTTATTACAAAATTCTCACTTCCTCTGATCTCTGAGCCGGCGG     540
gi|126033153|gb|EF076731.1|       AGGAAGGGGGCTGCTAGGTTATTACAAAATTCTCACTTCCTCTGATCTCTGAGCCGGCGG     540
gi|126033145|gb|EF076727.1|       AGGAAGGGGGCTGCTAGGTTATTACAAAATTCTCACTTCCTCTGATCTCTGAGCCGGCGG     540
gi|126033141|gb|EF076725.1|       AGGAAGGGGGCTGCTAGGTTATTACAAAATTCTCACTTCCTCTGATCTCTGAGCCGGCGG     540
gi|126033139|gb|EF076724.1|       AGGAAGGGGGCTGCTAGGTTATTACAAAATTCTCACTTCCTCTGATCTCTGAGCCGGCGG     540
gi|126033149|gb|EF076729.1|       AGGAAGGGGGCTGCTAGGTTATTACAAAATTCTCACTTCCTCTGATCTCTGAGCCGGCGG     540
gi|126033147|gb|EF076728.1|       AGGAAGGGGGCTGCTAGGTTATTACAAAATTCTCACTTCCTCTGATCTCTGAGCCGGCGG     540
gi|126033137|gb|EF076723.1|       AGGAAGGGGGCTGCTAGGTTATTACAAAATTCTCACTTCCTCTGATCTCTGAGCCGGCGG     540
gi|57471257|gb|AY859726.1|        ------------------------------------------------------------     0
gi|27848078|emb|AJ509824.2|       AGGAAGGGGGCTGCTAGGTTATTACAAAATTCTCACTTCCTCTGATCTCTGAGCCGGCGG     267
gi|34147166|ref|NM_183081.1|      ------------------------------------------------------------     0
gi|23821159|emb|AJ509825.1|       ------------------------------------------------------------     0


gi|407316560|gb|FJ495080.2|       CACGGGAAGTGGGCGCCAAGCATCCTTCCCTGCAGCTGCCTCCCAACCTGCCCGCCAGAC     585
```

```
gi|126033143|gb|EF076726.1|       CACGGGAAGTGGGCGCCAAGCATCCTTCCCTGCAGCTGCCTCCCAACCTGCCCGCCAGAC     600
gi|126033153|gb|EF076731.1|       CACGGGAAGTGGGCGCCAAGCATCCTTCCCTGCAGCTGCCTCCCAACCTGCCCGCCAGAC     600
gi|126033145|gb|EF076727.1|       CACGGGAAGTGGGCGCCAAGCATCCTTCCCTGCAGCTGCCTCCCAACCTGCCCGCCAGAC     600
gi|126033141|gb|EF076725.1|       CACGGGAAGTGGGCGCCAAGCATCCTTCCCTGCAGCTGCCTCCCAACCTGCCCGCCAGAC     600
gi|126033139|gb|EF076724.1|       CACGGGAAGTGGGCGCCAAGCATCCTTCCCTGCAGCTGCCTCCCAACCTGCCCGCCAGAC     600
gi|126033149|gb|EF076729.1|       CACGGGAAGTGGGCGCCAAGCATCCTTCCCTGCAGCTGCCTCCCAACCTGCCCGCCAGAC     600
gi|126033147|gb|EF076728.1|       CACGGGAAGTGGGCGCCAAGCATCCTTCCCTGCAGCTGCCTCCCAACCTGCCCGCCAGAC     600
gi|126033137|gb|EF076723.1|       CACGGGAAGTGGGCGCCAAGCATCCTTCCCTGCAGCTGCCTCCCAACCTGCCCGCCAGAC     600
gi|57471257|gb|AY859726.1|        ------------------------------------------------------------     0
gi|27848078|emb|AJ509824.2|       CACGGGAAGTGGGCGCCAAGCATCCTTCCCTGCAGCTGCCTCCCAACCTGCCCGCCAGAC     327
gi|34147166|ref|NM_183081.1|      ------------------------------------------------------------     0
gi|23821159|emb|AJ509825.1|       ------------------------------------------------------------     0

gi|407316560|gb|FJ495080.2|       CCTCTGGAGAAGCCGCATTCCCTGTCATGGTAGGATAGCAGCCCTCGGCCCCCGGGGAAG     645
gi|126033143|gb|EF076726.1|       CCTCTGGAGAAGCCGCATTCCCTGTCATGGTAGGATAGCAGCCCTCGGCCCCCGGGGAAG     660
gi|126033153|gb|EF076731.1|       CCTCTGGAGAAGCCGCATTCCCTGTCATGGTAGGATAGCAGCCCTCGGCCCCCGGGGAAG     660
gi|126033145|gb|EF076727.1|       CCTCTGGAGAAGCCGCATTCCCTGTCATGGTAGGATAGCAGCCCTCGGCCCCCGGGGAAG     660
gi|126033141|gb|EF076725.1|       CCTCTGGAGAAGCCGCATTCCCTGTCATGGTAGGATAGCAGCCCTCGGCCCCCGGGGAAG     660
gi|126033139|gb|EF076724.1|       CCTCTGGAGAAGCCGCATTCCCTGTCATGGTAGGATAGCAGCCCTCGGCCCCCGGGGAAG     660
gi|126033149|gb|EF076729.1|       CCTCTGGAGAAGCCGCATTCCCTGTCATGGTAGGATAGCAGCCCTCGGCCCCCGGGGAAG     660
gi|126033147|gb|EF076728.1|       CCTCTGGAGAAGCCGCATTCCCTGTCATGGTAGGATAGCAGCCCTCGGCCCCCGGGGAAG     660
gi|126033137|gb|EF076723.1|       CCTCTGGAGAAGCCGCATTCCCTGTCATGGTAGGATAGCAGCCCTCGGCCCCCGGGGAAG     660
gi|57471257|gb|AY859726.1|        ------------------------------------------------------------     0
gi|27848078|emb|AJ509824.2|       CCTCTGGAGAAGCCGCATTCCCTGTCATGGTAGGATAGCAGCCCTCGGCCCCCGGGGAAG     387
gi|34147166|ref|NM_183081.1|      ------------------------------------------------------------     0
gi|23821159|emb|AJ509825.1|       ------------------------------------------------------------     0

gi|407316560|gb|FJ495080.2|       AGGACAAGGTGGGGAGGAAGTGATGGAGAGAAACACGGGCTTGGGAGCTACAGACAAGGG     705
gi|126033143|gb|EF076726.1|       AGGACAAGGTGGGGAGGAAGTGATGGAGAGAAACACGGGCTTGGGAGCTACAGACAAGGG     720
gi|126033153|gb|EF076731.1|       AGGACAAGGTGGGGAGGAAGTGATGGAGAGAAACACGGGCTTGGGAGCTACAGACAAGGG     720
gi|126033145|gb|EF076727.1|       AGGACAAGGTGGGGAGGAAGTGATGGAGAGAAACACGGGCTTGGGAGCTACAGACAAGGG     720
gi|126033141|gb|EF076725.1|       AGGACAAGGTGGGGAGGAAGTGATGGAGAGAAACACGGGCTTGGGAGCTACAGACAAGGG     720
gi|126033139|gb|EF076724.1|       AGGACAAGGTGGGGAGGAAGTGATGGAGAGAAACACGGGCTTGGGAGCTACAGACAAGGG     720
gi|126033149|gb|EF076729.1|       AGGACAAGGTGGGGAGGAAGTGATGGAGAGAAACACGGGCTTGGGAGCTACAGACAAGGG     720
gi|126033147|gb|EF076728.1|       AGGACAAGGTGGGGAGGAAGTGATGGAGAGAAACACGGGCTTGGGAGCTACAGACAAGGG     720
gi|126033137|gb|EF076723.1|       AGGACAAGGTGGGGAGGAAGTGATGGAGAGAAACACGGGCTTGGGAGCTACAGACAAGGG     720
gi|57471257|gb|AY859726.1|        ------------------------------------------------------------     0
```

```
gi|27848078|emb|AJ509824.2|       AGGACAAGGTGGGGAGGAAGTGATGGAGAGAAACACGGGCTTGGGAGCTACAGACAAGGG     447
gi|34147166|ref|NM_183081.1|      ------------------------------------------------------------     0
gi|23821159|emb|AJ509825.1|       ------------------------------------------------------------     0


gi|407316560|gb|FJ495080.2|       TGGGTGAAAGGCACAGTCACCCAGGTGGCTCTGGGGAAATGCACTGAGCCCTGAAGTCAT     765
gi|126033143|gb|EF076726.1|       TGGGTGAAAGGCACAGTCACCCAGGTGGCTCTGGGGAAATGCACTGAGCCCTGAAGTCAT     780
gi|126033153|gb|EF076731.1|       TGGGTGAAAGGCACAGTCACCCAGGTGGCTCTGGGGAAATGCACTGAGCCCTGAAGTCAT     780
gi|126033145|gb|EF076727.1|       TGGGTGAAAGGCACAGTCACCCAGGTGGCTCTGGGGAAATGCACTGAGCCCTGAAGTCAT     780
gi|126033141|gb|EF076725.1|       TGGGTGAAAGGCACAGTCACCCAGGTGGCTCTGGGGAAATGCACTGAGCCCTGAAGTCAT     780
gi|126033139|gb|EF076724.1|       TGGGTGAAAGGCACAGTCACCCAGGTGGCTCTGGGGAAATGCACTGAGCCCTGAAGTCAT     780
gi|126033149|gb|EF076729.1|       TGGGTGAAAGGCACAGTCACCCAGGTGGCTCTGGGGAAATGCACTGAGCCCTGAAGTCAT     780
gi|126033147|gb|EF076728.1|       TGGGTGAAAGGCACAGTCACCCAGGTGGCTCTGGGGAAATGCACTGAGCCCTGAAGTCAT     780
gi|126033137|gb|EF076723.1|       TGGGTGAAAGGCACAGTCACCCAGGTGGCTCTGGGGAAATGCACTGAGCCCTGAAGTCAT     780
gi|57471257|gb|AY859726.1|        ------------------------------------------------------------     0
gi|27848078|emb|AJ509824.2|       TGGGTGAAAGGCACAGTCACCCAGGTGGCTCTGGGGAAATGCACTGAGCCCTGAAGTCAT     507
gi|34147166|ref|NM_183081.1|      ------------------------------------------------------------     0
gi|23821159|emb|AJ509825.1|       ------------------------------------------------------------     0


gi|407316560|gb|FJ495080.2|       GAGCCATGAGACCCAAACTCTGCTTGGGTGAGGAGGCTGGGGCTAAGGGATGCTCTCAGT     825
gi|126033143|gb|EF076726.1|       GAGCCATGAGACCCAAACTCTGCTTGGGTGAGGAGGCTGGGGCTAAGGGATGCTCTCAGT     840
gi|126033153|gb|EF076731.1|       GAGCCATGAGACCCAAACTCTGCTTGGGTGAGGAGGCTGGGGCTAAGGGATGCTCTCAGT     840
gi|126033145|gb|EF076727.1|       GAGCCATGAGACCCAAACTCTGCTTGGGTGAGGAGGCTGGGGCTAAGGGATGCTCTCAGT     840
gi|126033141|gb|EF076725.1|       GAGCCATGAGACCCAAACTCTGCTTGGGTGAGGAGGCTGGGGCTAAGGGATGCTCTCAGT     840
gi|126033139|gb|EF076724.1|       GAGCCATGAGACCCAAACTCTGCTTGGGTGAGGAGGCTGGGGCTAAGGGATGCTCTCAGT     840
gi|126033149|gb|EF076729.1|       GAGCCATGAGACCCAAACTCTGCTTGGGTGAGGAGGCTGGGGCTAAGGGATGCTCTCAGT     840
gi|126033147|gb|EF076728.1|       GAGCCATGAGACCCAAACTCTGCTTGGGTGAGGAGGCTGGGGCTAAGGGATGCTCTCAGT     840
gi|126033137|gb|EF076723.1|       GAGCCATGAGACCCAAACTCTGCTTGGGTGAGGAGGCTGGGGCTAAGGGATGCTCTCAGT     840
gi|57471257|gb|AY859726.1|        ------------------------------------------------------------     0
gi|27848078|emb|AJ509824.2|       GAGCCATGATACCCAAACTCTGCTTGGGTGAGGAGGCTGGGGCTAAGGGATGCTCTCAGT     567
gi|34147166|ref|NM_183081.1|      ------------------------------------------------------------     0
gi|23821159|emb|AJ509825.1|       ------------------------------------------------------------     0


gi|407316560|gb|FJ495080.2|       GGTGGTGCTGGGCTCACTGGTGGTGCTGGGCTCACTGGTGGTGGTGCTGGGCTCACTGGT     885
gi|126033143|gb|EF076726.1|       GGTGGTGCTGGGCTCACTGGTGGTGCTGGGCTCACTGGTGGTGSTGCTGGGCTCACTGGT     900
gi|126033153|gb|EF076731.1|       GGTGGTGCTGGGCTCACTGGTGGTGCTGGGCTCACTGGTGGTGGTGCTGGGCTCACTGGT     900
gi|126033145|gb|EF076727.1|       GGTGGTGCTGGGCTCACTGGTGGTGCTGGGCTCACTGGTGGTGGTGCTGGGCTCACTGGT     900
```

```
gi|126033141|gb|EF076725.1|      GGTGGTGCTGGGCTCACTGGTGGTGCTGGGCTCACTGGTGGTGGTGCTGGGCTCACTGGT     900
gi|126033139|gb|EF076724.1|      GGTGGTGCTGGGCTCACTGGTGGTGCTGGGCTCACTGGTGGTGGTGCTGGGCTCACTGGT     900
gi|126033149|gb|EF076729.1|      GGTGGTGCTGGGCTCACTGGTGGTGCTGGGCTCACTGGTGGTGGTGCTGGGCTCACTGGT     900
gi|126033147|gb|EF076728.1|      GGTGGTGCTGGGCTCACTGGTGGTGCTGGGCTCACTGGTGGTGGTGCTGGGCTCACTGGT     900
gi|126033137|gb|EF076723.1|      GGTGGTGCTGGGCTCACTGGTGGTGCTGGGCTCACTGGTGGTGGTGCTGGGCTCACTGGT     900
gi|57471257|gb|AY859726.1|       ------------------------------------------------------------     0
gi|27848078|emb|AJ509824.2|      GGTGGTGCTGGGCTCACTGGTGGTGCTGGGCTCACTGGTGGTGGTGCTGGGCTCACTGGT     627
gi|34147166|ref|NM_183081.1|     ------------------------------------------------------------     0
gi|23821159|emb|AJ509825.1|      ------------------------------------------------------------     0

gi|407316560|gb|FJ495080.2|      GGTGGTGGTGCTGGGCTCACTGGTGGTGGTGGTGCTGGGCTGCAAGCCAAGGGGATGCCA     945
gi|126033143|gb|EF076726.1|      GGTGGTGGTGCTGGGCTCACTGGTGGTGGTGGTGCTGGGCTGCAAGCCAAGGGGATGCCA     960
gi|126033153|gb|EF076731.1|      GGTGGTGGTGCTGGGCTCACTGGTGGTGGTGGTGCTGGGCTGCAAGCCAAGGGGACGCCA     960
gi|126033145|gb|EF076727.1|      GGTGGTGGTGCTGGGCTCACTGGTGGTGGTGGTGCTGGGCTGCAAGCCAAGGGGACGCCA     960
gi|126033141|gb|EF076725.1|      GGTGGTGGTGCTGGGCTCACTGGTGGTGGTGGTGCTGGGCTGCAAGCCAAGGGGACGCCA     960
gi|126033139|gb|EF076724.1|      GGTGGTGGTGCTGGGCTCACTGGTGGTGGTGGTGCTGGGCTGCAAGCCAAGGGGACGCCA     960
gi|126033149|gb|EF076729.1|      GGTGGTGGTGCTGGGCTCACTGGTGGTGGTGGTGCTGGGCTGCAAGCCAAGGGGATGCCA     960
gi|126033147|gb|EF076728.1|      GGTGGTGGTGCTGGGCTCACTGGTGGTGGTGGTGCTGGGCTGCAAGCCAAGGGGATGCCA     960
gi|126033137|gb|EF076723.1|      GGTGGTGGTGCTGGGCTCACTGGTGGTGGTGGTGCTGGGCTGCAAGCCAAGGGGATGCCA     960
gi|57471257|gb|AY859726.1|       ------------------------------------------------------------     0
gi|27848078|emb|AJ509824.2|      GGTGGTGGTGCTGGGCTCACTGGTGGTGGTGGTGCTGGGCTGCATGCCAAGGGGATGCCA     687
gi|34147166|ref|NM_183081.1|     ------------------------------------------------------------     0
gi|23821159|emb|AJ509825.1|      ------------------------------------------------------------     0

gi|407316560|gb|FJ495080.2|      GGGCTATGCCCTGTCTGCCCTTTGCTGCCCAACCTTCCCTCCAATTCGTTCTGACCCACA     1005
gi|126033143|gb|EF076726.1|      GGGCTATGCCCTGTCTGCCCTTTGCTGCCCAACCTTCCCTCCAATTCGTTCTGACCCACA     1020
gi|126033153|gb|EF076731.1|      GGGCTATGCCCTGTCTGCCCTTTGCTGCCCAACCTTCCCTCCAATTCGTTCTGACCCACA     1020
gi|126033145|gb|EF076727.1|      GGGCTATGCCCTGTCTGCCCTTTGCTGCCCAACCTTCCCTCCAATTCGTTCTGACCCACA     1020
gi|126033141|gb|EF076725.1|      GGGCTATGCCCTGTCTGCCCTTTGCTGCCCAACCTTCCCTCCAATTCGTTCTGACCCACA     1020
gi|126033139|gb|EF076724.1|      GGGCTATGCCCTGTCTGCCCTTTGCTGCCCAACCTTCCCTCCAATTCGTTCTGACCCACA     1020
gi|126033149|gb|EF076729.1|      GGGCTATGCCCTGTCTGCCCTTTGCTGCCCAACCTTCCCTCCAATTCGTTCTGACCCACA     1020
gi|126033147|gb|EF076728.1|      GGGCTATGCCCTGTCTGCCCTTTGCTGCCCAACCTTCCCTCCAATTCGTTCTGACCCACA     1020
gi|126033137|gb|EF076723.1|      GGGCTATGCCCTGTCTGCCCTTTGCTGCCCAACCTTCCCTCCAATTCGTTCTGACCCACA     1020
gi|57471257|gb|AY859726.1|       ------------------------------------------------------------     0
gi|27848078|emb|AJ509824.2|      GGGCTATGCCCTGTCTGCCCTTTGCTGCCCAACCTTCCCTCCAATTTGTTCTGACCCACA     747
gi|34147166|ref|NM_183081.1|     ------------------------------------------------------------     0
gi|23821159|emb|AJ509825.1|      ------------------------------------------------------------     0
```

```
gi|407316560|gb|FJ495080.2|       GCAGT-TTGAGTGCCTGGGCTGGCTAAGGTGGGCAGAGCATGAAAGAGCTAGGCAGTGGC     1064
gi|126033143|gb|EF076726.1|       GCAGT-TTGAGTGCCTGGGCTGGCTAAGGTGGGCAGAGCATGAAAGAGCTAGGCAGTGGC     1079
gi|126033153|gb|EF076731.1|       GCAGT-TTGAGTGCCTGGGCTGGCTAAGGTGGGCAGAGCATGAAAGAGCTAGGCAGTGGC     1079
gi|126033145|gb|EF076727.1|       GCAGT-TTGAGTGCCTGGGCTGGCTAAGGTGGGCAGAGCATGAAAGAGCTAGGCAGTGGC     1079
gi|126033141|gb|EF076725.1|       GCAGT-TTGAGTGCCTGGGCTGGCTAAGGTGGGCAGAGCATGAAAGAGCTAGGCAGTGGC     1079
gi|126033139|gb|EF076724.1|       GCAGT-TTGAGTGCCTGGGCTGGCTAAGGTGGGCAGAGCATGAAAGAGCTAGGCAGTGGC     1079
gi|126033149|gb|EF076729.1|       GCAGT-TTGAGTGCCTGGGCTGGCTAAGGTGGGCAGAGCATGAAAGAGCTAGGCAGTGGC     1079
gi|126033147|gb|EF076728.1|       GCAGT-TTGAGTGCCTGGGCTGGCTAAGGTGGGCAGAGCATGAAAGAGCTAGGCAGTGGC     1079
gi|126033137|gb|EF076723.1|       GCAGT-TTGAGTGCCTGGGCTGGCTAAGGTGGGCAGAGCATGAAAGAGCTAGGCAGTGGC     1079
gi|57471257|gb|AY859726.1|        ------------------------------------------------------------     0
gi|27848078|emb|AJ509824.2|       GCAGTTTTGAGTGCCTGGGCTGGCTAAGGTGGACAGAGCATGAAAGAGCTAGGCAGTGGC     807
gi|34147166|ref|NM_183081.1|      ------------------------------------------------------------     0
gi|23821159|emb|AJ509825.1|       ------------------------------------------------------------     0

gi|407316560|gb|FJ495080.2|       CATCCTGGCCAAGGCCCTGGGCAGTGGCCAGCTTGAACCCCAGCAGGAGACCTTCACTCC     1124
gi|126033143|gb|EF076726.1|       CATCCTGGCCAAGGCCCTGGGCAGTGGCCAGCTTGAACCCCAGCAGGAGACCTTCACTCC     1139
gi|126033153|gb|EF076731.1|       CATCCTGGCCAAGGCCCTGGGCAGTGGCCAGCTTGAACCCCAGCAGGAGACCTTCACTCC     1139
gi|126033145|gb|EF076727.1|       CATCCTGGCCAAGGCCCTGGGCAGTGGCCAGCTTGAACCCCAGCAGGAGACCTTCACTCC     1139
gi|126033141|gb|EF076725.1|       CATCCTGGCCAAGGCCCTGGGCAGTGGCCAGCTTGAACCCCAGCAGGAGACCTTCACTCC     1139
gi|126033139|gb|EF076724.1|       CATCCTGGCCAAGGCCCTGGGCAGTGGCCAGCTTGAACCCCAGCAGGAGACCTTCACTCC     1139
gi|126033149|gb|EF076729.1|       CATCCTGGCCAAGGCCCTGGGCAGTGGCCAGCTTGAACCCCAGCAGGAGACCTTCACTCC     1139
gi|126033147|gb|EF076728.1|       CATCCTGGCCAAGGCCCTGGGCAGTGGCCAGCTTGAACCCCAGCAGGAGACCTTCACTCC     1139
gi|126033137|gb|EF076723.1|       CATCCTGGCCAAGGCCCTGGGCAGTGGCCAGCTTGAACCCCAGCAGGAGACCTTCACTCC     1139
gi|57471257|gb|AY859726.1|        ------------------------------------------------------------     0
gi|27848078|emb|AJ509824.2|       CATCCTGTCCAAGGCCCTGGGCAGTGGCCAGCTTGAACCCCAGCAGGAGACCTTCACTCC     867
gi|34147166|ref|NM_183081.1|      ------------------------------------------------------------     0
gi|23821159|emb|AJ509825.1|       ------------------------------------------------------------     0

gi|407316560|gb|FJ495080.2|       CACCTCTTGGCCAAACGGAGCAGGAGTGGTTGGGGGGAGCCTGACAGACAGCAGAGGGAG     1184
gi|126033143|gb|EF076726.1|       CACCTCTTGGCCAAACGGAGCAGGAGTGGTTGGGGGGAGCCTGACAGACAGCAGAGGGAG     1199
gi|126033153|gb|EF076731.1|       CACCTCTTGGCCAAACGGAGCAGGAGTGGTTGGGGGGAGCCTGACAGACAGCAGAGGGAG     1199
gi|126033145|gb|EF076727.1|       CACCTCTTGGCCAAACGGAGCAGGAGTGGTTGGGGGGAGCCTGACAGACAGCAGAGGGAG     1199
gi|126033141|gb|EF076725.1|       CACCTCTTGGCCAAACGGAGCAGGAGTGGTTGGGGGGAGCCTGACAGACAGCAGAGGGAG     1199
gi|126033139|gb|EF076724.1|       CACCTCTTGGCCAAACGGAGCAGGAGTGGTTGGGGGGAGCCTGACAGACAGCAGAGGGAG     1199
gi|126033149|gb|EF076729.1|       CACCTCTTGGCCAAACGGAGCAGGAGTGGTTGGGGGGAGCCTGACAGACAGCAGAGGGAG     1199
```

```
gi|126033147|gb|EF076728.1|       CACCTCTTGGCCAAACGGAGCAGGAGTGGTTGGGGGGAGCCTGACAGACAGCAGAGGGAG     1199
gi|126033137|gb|EF076723.1|       CACCTCTTGGCCAAACGGAGCAGGAGTGGTTGGGGGGAGCCTGACAGACAGCAGAGGGAG     1199
gi|57471257|gb|AY859726.1|        ------------------------------------------------------------     0
gi|27848078|emb|AJ509824.2|       CACCTCTTGGCCAAACGGAGCAGGAGTGGTTGGGGGGAGCCTGACAGACAGCAGAGGGAG     927
gi|34147166|ref|NM_183081.1|      ------------------------------------------------------------     0
gi|23821159|emb|AJ509825.1|       ------------------------------------------------------------     0


gi|407316560|gb|FJ495080.2|       GTTTTCAGGGGTGCTGCAGGGAAGGGAAGATGCCTCTTTGCTTGGGAGAGACGAGAAGTG     1244
gi|126033143|gb|EF076726.1|       GTTTTCAGGGGTGCTGCAGGGAAGGGAAGATGCCTCTTTGCTTGGGAGAGACGAGAAGTG     1259
gi|126033153|gb|EF076731.1|       GTTTTCAGGGGTGCTGCAGGGAAGGGAAGATGCCTCTTTGCTTGGGAGAGACGAGAAGTG     1259
gi|126033145|gb|EF076727.1|       GTTTTCAGGGGTGCTGCAGGGAAGGGAAGATGCCTCTTTGCTTGGGAGAGACGAGAAGTG     1259
gi|126033141|gb|EF076725.1|       GTTTTCAGGGGTGCTGCAGGGAAGGGAAGATGCCTCTTTGCTTGGGAGAGACGAGAAGTG     1259
gi|126033139|gb|EF076724.1|       GTTTTCAGGGGTGCTGCAGGGAAGGGAAGATGCCTCTTTGCTTGGGAGAGACGAGAAGTG     1259
gi|126033149|gb|EF076729.1|       GTTTTCAGGGGTGCTGCAGGGAAGGGAAGATGCCTCTTTGCTTGGGAGAGACGAGAAGTG     1259
gi|126033147|gb|EF076728.1|       GTTTTCAGGGGTGCTGCAGGGAAGGGAAGATGCCTCTTTGCTTGGGAGAGACGAGAAGTG     1259
gi|126033137|gb|EF076723.1|       GTTTTCAGGGGTGCTGCAGGGAAGGGAAGATGCCTCTTTGCTTGGGAGAGACGAGAAGTG     1259
gi|57471257|gb|AY859726.1|        ------------------------------------------------------------     0
gi|27848078|emb|AJ509824.2|       GTTTTCAGGGGTGCTGCAGGGAAGGGAAGATGCCTCTTTGCTTGGGAGAGACGAGAAGTG     987
gi|34147166|ref|NM_183081.1|      ------------------------------------------------------------     0
gi|23821159|emb|AJ509825.1|       ------------------------------------------------------------     0


gi|407316560|gb|FJ495080.2|       ACCTTCTCGTCTCTCCCAAGAGAAGGAATGGAAGGGGGAAGCGAGGATGAGCCTGGGGGG     1304
gi|126033143|gb|EF076726.1|       ACCTTCTCGTCTCTCCCAAGAGAAGGAATGGAAGGGGGAAGCGAGGATGAGCCTGGGGGG     1319
gi|126033153|gb|EF076731.1|       ACCTTCTCGTCTCTCCCAAGAGAAGGAATGGAAGGGGGAAGCGAGGATGAGCCTGGGGGG     1319
gi|126033145|gb|EF076727.1|       ACCTTCTCGTCTCTCCCAAGAGAAGGAATGGAAGGGGGAAGCGAGGATGAGCCTGGGGGG     1319
gi|126033141|gb|EF076725.1|       ACCTTCTCGTCTCTCCCAAGAGAAGGAATGGAAGGGGGAAGCGAGGATGAGCCTGGGGGG     1319
gi|126033139|gb|EF076724.1|       ACCTTCTCGTCTCTCCCAAGAGAAGGAATGGAAGGGGGAAGCGAGGATGAGCCTGGGGGG     1319
gi|126033149|gb|EF076729.1|       ACCTTCTCGTCTCTCCCAAGAGAAGGAATGGAAGGGGGAAGCGAGGATGAGCCTGGGGGG     1319
gi|126033147|gb|EF076728.1|       ACCTTCTCGTCTCTCCCAAGAGAAGGAATGGAAGGGGGAAGCGAGGATGAGCCTGGGGGG     1319
gi|126033137|gb|EF076723.1|       ACCTTCTCGTCTCTCCCAAGAGAAGGAATGGAAGGGGGAAGCGAGGATGAGCCTGGGGGG     1319
gi|57471257|gb|AY859726.1|        ------------------------------------------------------------     0
gi|27848078|emb|AJ509824.2|       ACCTTCTCGTCTCTCCCAAGAGAAGGAATGGAAGGGGGAAGCGAGGATGAGCCTGGGGGG     1047
gi|34147166|ref|NM_183081.1|      ------------------------------------------------------------     0
gi|23821159|emb|AJ509825.1|       ------------------------------------------------------------     0


gi|407316560|gb|FJ495080.2|       CCTGTTGACCACCAACCTGGCTCTCCTGTTCAGCTAGGGTTTGGTTTCGAGTGAAAGCCT     1364
```

```
gi|126033143|gb|EF076726.1|       CCTGTTGACCACCAACCTGGCTCTCCTGTCCAGCTAGGGTTTGGTTTCGAGTGAAAGCCT     1379
gi|126033153|gb|EF076731.1|       CCTGTTGACCACCAACCTGGCTCTCCTGTTCAGCTAGGGTTTGGTTTCGAGTGAAAGCCT     1379
gi|126033145|gb|EF076727.1|       CCTGTTGACCACCAACCTGGCTCTCCTGTTCAGCTAGGGTTTGGTTTCGAGTGAAAGCCT     1379
gi|126033141|gb|EF076725.1|       CCTGTTGACCACCAACCTGGCTCTCCTGTTCAGCTAGGGTTTGGTTTCGAGTGAAAGCCT     1379
gi|126033139|gb|EF076724.1|       CCTGTTGACCACCAACCTGGCTCTCCTGTTCAGCTAGGGTTTGGTTTCGAGTGAAAGCCT     1379
gi|126033149|gb|EF076729.1|       CCTGTTGACCACCAACCTGGCTCTCCTGTTCAGCTAGGGTTTGGTTTCGAGTGAAAGCCT     1379
gi|126033147|gb|EF076728.1|       CCTGTTGACCACCAACCTGGCTCTCCTGTTCAGCTAGGGTTTGGTTTCGAGTGAAAGCCT     1379
gi|126033137|gb|EF076723.1|       CCTGTTGACCACCAACCTGGCTCTCCTGTTCAGCTAGGGTTTGGTTTCGAGTGAAAGCCT     1379
gi|57471257|gb|AY859726.1|        ------------------------------------------------------------     0
gi|27848078|emb|AJ509824.2|       CCTGTTGACCACCAACCTGGCTCTCCTGTCCAGCTAGGGTTTGGTTTCGAGTGAAAGCCT     1107
gi|34147166|ref|NM_183081.1|      ------------------------------------------------------------     0
gi|23821159|emb|AJ509825.1|       ------------------------------------------------------------     0

gi|407316560|gb|FJ495080.2|       CCATCACGAGTCTAGATTGCAGGTCTCAGGATGAGCTCTGATTCCTTTCTCTGAGTCAAA     1424
gi|126033143|gb|EF076726.1|       CCATCACGAGTCTAGATTGCAGGTCTCAGGATGAGCTCTGATTCCTTTCTCTGAGTCAAA     1439
gi|126033153|gb|EF076731.1|       CCATCACGAGTCTAGATTGCAGGTCTCAGGATGAGCTCTGATTCCTTTCTCTGAGTCAAA     1439
gi|126033145|gb|EF076727.1|       CCATCACGAGTCTAGATTGCAGGTCTCAGGATGAGCTCTGATTCCTTTCTCTGAGTCAAA     1439
gi|126033141|gb|EF076725.1|       CCATCACGAGTCTAGATTGCAGGTCTCAGGATGAGCTCTGATTCCTTTCTCTGAGTCAAA     1439
gi|126033139|gb|EF076724.1|       CCATCACGAGTCTAGATTGCAGGTCTCAGGATGAGCTCTGATTCCTTTCTCTGAGTCAAA     1439
gi|126033149|gb|EF076729.1|       CCATCACGAGTCTAGATTGCAGGTCTCAGGATGAGCTCTGATTCCTTTCTCTGAGTCAAA     1439
gi|126033147|gb|EF076728.1|       CCATCACGAGTCTAGATTGCAGGTCTCAGGATGAGCTCTGATTCCTTTCTCTGAGTCAAA     1439
gi|126033137|gb|EF076723.1|       CCATCACGAGTCTAGATTGCAGGTCTCAGGATGAGCTCTGATTCCTTTCTCTGAGTCAAA     1439
gi|57471257|gb|AY859726.1|        ------------------------------------------------------------     0
gi|27848078|emb|AJ509824.2|       CCATCACGAGTCTAGATTGCAAGTCTCAGGATGAGCTCTGATTCCTTTCTCCGAGTCAAA     1167
gi|34147166|ref|NM_183081.1|      ------------------------------------------------------------     0
gi|23821159|emb|AJ509825.1|       ------------------------------------------------------------     0

gi|407316560|gb|FJ495080.2|       GGCTGGGTCTCAGATCTGAGTCCAAGTTCCCCATGGGCCTTGGAGATGTGATGGGCACTG     1484
gi|126033143|gb|EF076726.1|       GGCTGGGTCTCAGATCTGAGTCCAAGTTCCCCATGGGCCTTGGAGATGTGATGGGCACTG     1499
gi|126033153|gb|EF076731.1|       GGCTGGGTCTCAGATCTGAGTCCAAGTTCCCCATGGGCCTTGGAGATGTGATGGGCACTG     1499
gi|126033145|gb|EF076727.1|       GGCTGGGTCTCAGATCTGAGTCCAAGTTCCCCATGGGCCTTGGAGATGTGATGGGCACTG     1499
gi|126033141|gb|EF076725.1|       GGCTGGGTCTCAGATCTGAGTCCAAGTTCCCCATGGGCCTTGGAGATGTGATGGGCACTG     1499
gi|126033139|gb|EF076724.1|       GGCTGGGTCTCAGATCTGAGTCCAAGTTCCCCATGGGCCTTGGAGATGTGATGGGCACTG     1499
gi|126033149|gb|EF076729.1|       GGCTGGGTCTCAGATCTGAGTCCAAGTTCCCCATGGGCCTTGGAGATGTGATGGGCACTG     1499
gi|126033147|gb|EF076728.1|       GGCTGGGTCTCAGATCTGAGTCCAAGTTCCCCATGGGCCTTGGAGATGTGATGGGCACTG     1499
gi|126033137|gb|EF076723.1|       GGCTGGGTCTCAGATCTGAGTCCAAGTTCCCCATGGGCCTTGGAGATGTGATGGGCACTG     1499
gi|57471257|gb|AY859726.1|        ------------------------------------------------------------     0
```

```
gi|27848078|emb|AJ509824.2|       GGCCGGGTCTCAGATCTGAGTCCAAGTTCCCCATGGGCCTTGGAGATGTGATGGGCACTG     1227
gi|34147166|ref|NM_183081.1|      ------------------------------------------------------------     0
gi|23821159|emb|AJ509825.1|       ------------------------------------------------------------     0


gi|407316560|gb|FJ495080.2|       GAGAGAGAAACAGGAACCTCCCTGGAGGGCTGGGAGAGACGGAGAGACAGAGGTGGAAGC     1544
gi|126033143|gb|EF076726.1|       GAGAGAGAAACAGGAACCTCCCTGGAGGGCTGGGAGAGACGGAGAGACAGAGGTGGAAGC     1559
gi|126033153|gb|EF076731.1|       GAGAGAGAAACAGGAACCTCCCTGGAGGGCTGGGAGAGACGGAGAGACAGAGGTGGAAGC     1559
gi|126033145|gb|EF076727.1|       GAGAGAGAAACAGGAACCTCCCTGGAGGGCTGGGAGAGACGGAGAGACAGAGGTGGAAGC     1559
gi|126033141|gb|EF076725.1|       GAGAGAGAAACAGGAACCTCCCTGGAGGGCTGGGAGAGACGGAGAGACAGAGGTGGAAGC     1559
gi|126033139|gb|EF076724.1|       GAGAGAGAAACAGGAACCTCCCTGGAGGGCTGGGAGAGACGGAGAGACAGAGGTGGAAGC     1559
gi|126033149|gb|EF076729.1|       GAGAGAGAAACAGGAACCTCCCTGGAGGGCTGGGAGAGACGGAGAGACAGAGGTGGAAGC     1559
gi|126033147|gb|EF076728.1|       GAGAGAGAAACAGGAACCTCCCTGGAGGGCTGGGAGAGACGGAGAGACAGAGGTGGAAGC     1559
gi|126033137|gb|EF076723.1|       GAGAGAGAAACAGGAACCTCCCTGGAGGGCTGGGAGAGACGGAGAGACAGAGGTGGAAGC     1559
gi|57471257|gb|AY859726.1|        ----------------------------------------------GGGAAGTGGGCGC     13
gi|27848078|emb|AJ509824.2|       GAGAGAGAAACAGGAACCTCCCTGGAGGGCTGGGAGAGACGGAGAGACAGAGGTGGAAGC     1287
gi|34147166|ref|NM_183081.1|      ----ACAAAATTCTCACTTCCTCTGATCTCTGAGCCGGCGGCACGGGAAGT---GGGCGC     53
gi|23821159|emb|AJ509825.1|       ------------------------------------------------------------     0


gi|407316560|gb|FJ495080.2|       CTTTCTGGGGAACCCCCAGGGAGGTGAGGCAGAGGAGGTGGAGCTGGGGCCGTGTTCTGA     1604
gi|126033143|gb|EF076726.1|       CCTTCTGGGGAACCCCCAGGGAGGTGAGGCAGAGGAGGTGGAGCTGGGGCCGTGTTCTGA     1619
gi|126033153|gb|EF076731.1|       CCTTCTGGGGAACCCCCAGGGAGGTGAGGCAGAGGAGGTGGAGCTGGGGCCGTGTTCTGA     1619
gi|126033145|gb|EF076727.1|       CCTTCTGGGGAACCCCCAGGGAGGTGAGGCAGAGGAGGTGGAGCTGGGGCCGTGTTCTGA     1619
gi|126033141|gb|EF076725.1|       CCTTCTGGGGAACCCCCAGGGAGGTGAGGCAGAGGAGGTGGAGCTGGGGCCGTGTTCTGA     1619
gi|126033139|gb|EF076724.1|       CCTTCTGGGGAACCCCCAGGGAGGTGAGGCAGAGGAGGTGGAGCTGGGGCCGTGTTCTGA     1619
gi|126033149|gb|EF076729.1|       CCTTCTGGGGAACCCCCAGGGAGGTGAGGCAGAGGAGGTGGAGCTGGGGCCGTGTTCTGA     1619
gi|126033147|gb|EF076728.1|       CCTTCTGGGGAACCCCCAGGGAGGTGAGGCAGAGGAGGTGGAGCTGGGGCCGTGTTCTGA     1619
gi|126033137|gb|EF076723.1|       CCTTCTGGGGAACCCCCAGGGAGGTGAGGCAGAGGAGGTGGAGCTGGGGCCGTGTTCTGA     1619
gi|57471257|gb|AY859726.1|        CAAG------CATCCTTCCCTGCAGCTGCCTCCCAACCTGCCCGCCAGACCCTCTGGAGA     67
gi|27848078|emb|AJ509824.2|       CCTTCTGGGGAACCCCCAGGGAGGTGAGGCGGAGGAGGTGGAGCTGGGGCCGTGTTCTGA     1347
gi|34147166|ref|NM_183081.1|      CAA------GCATCCTTCCCTGCAGCTGCCTCCCAACCTGCCCGCCAGACCCTCTGGAGA     107
gi|23821159|emb|AJ509825.1|       ------------------------------------------------------------     0


gi|407316560|gb|FJ495080.2|       CCCGGCGCCCACCTTGCAGGGCCCCTACTGTGCCCCGCACCCCCTTTCTCTCCTGGTGCA     1664
gi|126033143|gb|EF076726.1|       CCCGGCGCCCACCTTGCAGGGCCCCTACTGTGCCCCGCACCCCCTTTCTCTCCTGGTGCA     1679
gi|126033153|gb|EF076731.1|       CCCGGCGCCCACCTTGCAGGGCCCCTACTGTGCCCCGCACCCCCTTTCTCTCCTGGTGCA     1679
gi|126033145|gb|EF076727.1|       CCCGGCGCCCACCTTGCAGGGCCCCTACTGTGCCCCGCACCCCCTTTCTCTCCTGGTGCA     1679
```

```
gi|126033141|gb|EF076725.1|       CCCGGCGCCCACCTTGCAGGGCCCCTACTGTGCCCCGCACCCCCTTTCTCTCCTGGTGCA     1679
gi|126033139|gb|EF076724.1|       CCCGGCGCCCACCTTGCAGGGCCCCTACTGTGCCCCGCACCCCCTTTCTCTCCTGGTGCA     1679
gi|126033149|gb|EF076729.1|       CCCGGCGCCCACCTTGCAGGGCCCCTACTGTGCCCCGCACCCCCTTTCTCTCCTGGTGCA     1679
gi|126033147|gb|EF076728.1|       CCCGGCGCCCACCTTGCAGGGCCCCTACTGTGCCCCGCACCCCCTTTCTCTCCTGGTGCA     1679
gi|126033137|gb|EF076723.1|       CCCGGCGCCCACCTTGCAGGGCCCCTACTGTGCCCCGCACCCCCTTTCTCTCCTGGTGCA     1679
gi|57471257|gb|AY859726.1|        AGCCGCATTCCCTGTCATGGGCCCCTACTGTGCCCCGCACCCCCTTTCTCTCCTGGTGCA     127
gi|27848078|emb|AJ509824.2|       CCCGGCGCCCACCTTGCAGGGCCCCTACTGTGCCCCGCACCCCCTTTCTCTCCTGGTGCA     1407
gi|34147166|ref|NM_183081.1|      AGCCGCATTCCCTGTCATGGGCCCCTACTGTGCCCCGCACCCCCTTTCTCTCCTGGTGCA     167
gi|23821159|emb|AJ509825.1|       ------------------------------------CACCCCCTTTCTCTCCTGGTGCA     23
                                                                        **********************

gi|407316560|gb|FJ495080.2|       GGCGGCGGCACTGGCAGCGGCCCTGGCCGAGGGCACCCTGCCTGCCTTCCTGCCCTGTGA     1724
gi|126033143|gb|EF076726.1|       GGCGGCGGCACTGGCAGCGGCCCTGGCCGAGGGCACCCTGCCTGCCTTCCTGCCCTGTGA     1739
gi|126033153|gb|EF076731.1|       GGCGGCGGCACTGGCAGCGGCCCTGGCCGAGGGCACCCTGCCTGCCTTCCTGCCCTGTGA     1739
gi|126033145|gb|EF076727.1|       GGCGGCGGCACTGGCAGCGGCCCTGGCCGAGGGCACCCTGCCTGCCTTCCTGCCCTGTGA     1739
gi|126033141|gb|EF076725.1|       GGCGGCGGCACTGGCAGCGGCCCTGGCCGAGGGCACCCTGCCTGCCTTCCTGCCCTGTGA     1739
gi|126033139|gb|EF076724.1|       GGCGGCGGCACTGGCAGCGGCCCTGGCCGAGGGCACCCTGCCTGCCTTCCTGCCCTGTGA     1739
gi|126033149|gb|EF076729.1|       GGCGGCGGCACTGGCAGCGGCCCTGGCCGAGGGCACCCTGCCTGCCTTCCTGCCCTGTGA     1739
gi|126033147|gb|EF076728.1|       GGCGGCGGCACTGGCAGCGGCCCTGGCCGAGGGCACCCTGCCTGCCTTCCTGCCCTGTGA     1739
gi|126033137|gb|EF076723.1|       GGCGGCGGCACTGGCAGCGGCCCTGGCCGAGGGCACCCTGCCTGCCTTCCTGCCCTGTGA     1739
gi|57471257|gb|AY859726.1|        GGCGGCGGCACTGGCAGCGGCCCTGGCCGAGGGCACCCTGCCTGCCTTCCTGCCCTGTGA     187
gi|27848078|emb|AJ509824.2|       GGCGGCGGCACTGGCAGTGGCCCTGGCCGAGGGCACCCTGCCTGCCTTCCTGCCCTGTGA     1467
gi|34147166|ref|NM_183081.1|      GGCGGCGGCACTGGCAGTGGCCCTGGCCGAGGGCACCCTGCCTGCCTTCCTGCCCTGTGA     227
gi|23821159|emb|AJ509825.1|       GGCGGCGGCACTGGCAGTGGCCCTGGCCGAGGGCACCCTGCCTGCCTTCCTGCCCTGTGA     83
                                  ***************** ****************************************

gi|407316560|gb|FJ495080.2|       GCTCCAGCCCCATGGTCAGGTGGACTGCAACTGGCTGTTCCTGAAGTCTGTGCCGCACTT     1784
gi|126033143|gb|EF076726.1|       GCTCCAGCCCCATGGTCAGGTGGACTGCAACTGGCTGTTCCTGAAGTCTGTGCCGCACTT     1799
gi|126033153|gb|EF076731.1|       GCTCCAGCCCCATGGTCAGGTGGACTGCAACTGGCTGTTCCTGAAGTCTGTGCCGCACTT     1799
gi|126033145|gb|EF076727.1|       GCTCCAGCCCCATGGTCAGGTGGACTGCAACTGGCTGTTCCTGAAGTCTGTGCCGCACTT     1799
gi|126033141|gb|EF076725.1|       GCTCCAGCCCCATGGTCAGGTGGACTGCAACTGGCTGTTCCTGAAGTCTGTGCCGCACTT     1799
gi|126033139|gb|EF076724.1|       GCTCCAGCCCCATGGTCAGGTGGACTGCAACTGGCTGTTCCTGAAGTCTGTGCCGCACTT     1799
gi|126033149|gb|EF076729.1|       GCTCCAGCCCCATGGTCAGGTGGACTGCAACTGGCTGTTCCTGAAGTCTGTGCCGCACTT     1799
gi|126033147|gb|EF076728.1|       GCTCCAGCCCCATGGTCAGGTGGACTGCAACTGGCTGTTCCTGAAGTCTGTGCCGCACTT     1799
gi|126033137|gb|EF076723.1|       GCTCCAGCCCCATGGTCAGGTGGACTGCAACTGGCTGTTCCTGAAGTCTGTGCCGCACTT     1799
gi|57471257|gb|AY859726.1|        GCTCCAGCCCCATGGTCAGGTGGACTGCAACTGGCTGTTCCTGAAGTCTGTGCCGCACTT     247
gi|27848078|emb|AJ509824.2|       GCTCCAGCCCCATGGTCAGGTGGACTGCAACTGGCTGTTCCTGAAGTCTGTGCCGCACTT     1527
gi|34147166|ref|NM_183081.1|      GCTCCAGCCCCATGGTCAGGTGGACTGCAACTGGCTGTTCCTGAAGTCTGTGCCGCACTT     287
gi|23821159|emb|AJ509825.1|       GCTCCAGCCCCATGGTCAGGTGGACTGCAACTGGCTGTTCCTGAAGTCTGTGCCGCACTT     143
```

```
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       TTCGGCTGGAGCCCCCCGGGCCAATGTCACCAGCCTCTCCTTAATCTCCAACCGCATCCA     1844
gi|126033143|gb|EF076726.1|       TTCGGCTGGAGCCCCCCGGGCCAATGTCACCAGCCTCTCCTTAATCTCCAACCGCATCCA     1859
gi|126033153|gb|EF076731.1|       TTCGGCTGGAGCCCCCCGGGCCAATGTCACCAGCCTCTCCTTAATCTCCAACCGCATCCA     1859
gi|126033145|gb|EF076727.1|       TTCGGCTGGAGCCCCCCGGGCCAATGTCACCAGCCTCTCCTTAATCTCCAACCGCATCCA     1859
gi|126033141|gb|EF076725.1|       TTCGGCTGGAGCCCCCCGGGCCAATGTCACCAGCCTCTCCTTAATCTCCAACCGCATCCA     1859
gi|126033139|gb|EF076724.1|       TTCGGCTGGAGCCCCCCGGGCCAATGTCACCAGCCTCTCCTTAATCTCCAACCGCATCCA     1859
gi|126033149|gb|EF076729.1|       TTCGGCTGGAGCCCCCCGGGCCAATGTCACCAGCCTCTCCTTAATCTCCAACCGCATCCA     1859
gi|126033147|gb|EF076728.1|       TTCGGCTGGAGCCCCCCGGGCCAATGTCACCAGCCTCTCCTTAATCTCCAACCGCATCCA     1859
gi|126033137|gb|EF076723.1|       TTCGGCTGGAGCCCCCCGGGCCAATGTCACCAGCCTCTCCTTAATCTCCAACCGCATCCA     1859
gi|57471257|gb|AY859726.1|        TTCGGCTGGAGCCCCCCGGGCCAATGTCACCAGCCTCTCCTTAATCTCCAACCGCATCCA     307
gi|27848078|emb|AJ509824.2|       TTCGGCTGGAGCCCCCCGGGCCAATGTCACCAGCCTCTCCTTAATCTCCAACCGCATCCA     1587
gi|34147166|ref|NM_183081.1|      TTCGGCTGGAGCCCCCCGGGCCAATGTCACCAGCCTCTCCTTAATCTCCAACCGCATCCA     347
gi|23821159|emb|AJ509825.1|       TTCGGCTGGAGCCCCCCGGGCCAATGTCACCAGCCTCTCCTTAATCTCCAACCGCATCCA     203
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       CCACTTGCATGACTCTGACTTCGTCCACCTGTCCAACCTGCGGGTCCTCAACCTCAAGTG     1904
gi|126033143|gb|EF076726.1|       CCACTTGCATGACTCTGACTTCGTCCACCTGTCCAACCTGCGGGTCCTCAACCTCAAGTG     1919
gi|126033153|gb|EF076731.1|       CCACTTGCATGACTCTGACTTCGTCCACCTGTCCAACCTGCGGGTCCTCAACCTCAAGTG     1919
gi|126033145|gb|EF076727.1|       CCACTTGCATGACTCTGACTTCGTCCACCTGTCCAACCTGCGGGTCCTCAACCTCAAGTG     1919
gi|126033141|gb|EF076725.1|       CCACTTGCATGACTCTGACTTCGTCCACCTGTCCAACCTGCGGGTCCTCAACCTCAAGTG     1919
gi|126033139|gb|EF076724.1|       CCACTTGCATGACTCTGACTTCGTCCACCTGTCCAACCTGCGGGTCCTCAACCTCAAGTG     1919
gi|126033149|gb|EF076729.1|       CCACTTGCATGACTCTGACTTCGTCCACCTGTCCAACCTGCGGGTCCTCAACCTCAAGTG     1919
gi|126033147|gb|EF076728.1|       CCACTTGCATGACTCTGACTTCGTCCACCTGTCCAACCTGCGGGTCCTCAACCTCAAGTG     1919
gi|126033137|gb|EF076723.1|       CCACTTGCATGACTCTGACTTCGTCCACCTGTCCAACCTGCGGGTCCTCAACCTCAAGTG     1919
gi|57471257|gb|AY859726.1|        CCACTTGCATGACTCTGACTTCGTCCACCTGTCCAACCTGCGGGTCCTCAACCTCAAGTG     367
gi|27848078|emb|AJ509824.2|       CCACTTGCATGACTCTGACTTCGCCCACCTGTCCAACCTGCGGGTCCTCAACCTCAAGTG     1647
gi|34147166|ref|NM_183081.1|      CCACTTGCATGACTCTGACTTCGCCCACCTGTCCAACCTGCGGGTCCTCAACCTCAAGTG     407
gi|23821159|emb|AJ509825.1|       CCACTTGCATGACTCTGACTTCGCCCACCTGTCCAACCTGCGGGTCCTCAACCTCAAGTG     263
                                  ********************** ***********************************

gi|407316560|gb|FJ495080.2|       GAACTGCCCGCCGGCYGGCCTCAGCCCCATGCACTTCCCCTGCCGTATGACCATCGAGCC     1964
gi|126033143|gb|EF076726.1|       GAACTGCCCGCCGGCCGGCCTCAGCCCCATGCACTTCCCCTGCCGTATGACCATCGAGCC     1979
gi|126033153|gb|EF076731.1|       GAACTGCCCGCCGGCCGGCCTCAGCCCCATGCACTTCCCCTGCCGTATGACCATCGAGCC     1979
gi|126033145|gb|EF076727.1|       GAACTGCCCGCCGGCCGGCCTCAGCCCCATGCACTTCCCCTGCCGTATGACCATCGAGCC     1979
gi|126033141|gb|EF076725.1|       GAACTGCCCGCCGGCCGGCCTCAGCCCCATGCACTTCCCCTGCCGTATGACCATCGAGCC     1979
gi|126033139|gb|EF076724.1|       GAACTGCCCGCCGGCCGGCCTCAGCCCCATGCACTTCCCCTGCCGTATGACCATCGAGCC     1979
gi|126033149|gb|EF076729.1|       GAACTGCCCGCCGGCCGGCCTCAGCCCCATGCACTTCCCCTGCCGTATGACCATCGAGCC     1979
```

```
gi|126033147|gb|EF076728.1|       GAACTGCCCGCCGGCCGGCCTCAGCCCCATGCACTTCCCCTGCCGTATGACCATCGAGCC     1979
gi|126033137|gb|EF076723.1|       GAACTGCCCGCCGGCCGGCCTCAGCCCCATGCACTTCCCCTGCCGTATGACCATCGAGCC     1979
gi|57471257|gb|AY859726.1|        GAACTGCCCGCCGGCCGGCCTCAGCCCCATGCACTTCCCCTGCCGTATGACCATCGAGCC     427
gi|27848078|emb|AJ509824.2|       GAACTGCCCGCCGGCCGGCCTCAGCCCCATGCACTTCCCCTGCCGTATGACCATCGAGCC     1707
gi|34147166|ref|NM_183081.1|      GAACTGCCCGCCGGCCGGCCTCAGCCCCATGCACTTCCCCTGCCGTATGACCATCGAGCC     467
gi|23821159|emb|AJ509825.1|       GAACTGCCCGCCGGCCGGCCTCAGCCCCATGCACTTCCCCTGCCGTATGACCATCGAGCC     323
                                  *************** ******************************************

gi|407316560|gb|FJ495080.2|       CAACACCTTCCTGGCTGTGCCCACCCTGGAGGAGCTGAACCTGAGCTACAACGGCATCAC     2024
gi|126033143|gb|EF076726.1|       CAACACCTTCCTGGCTGTGCCCACCCTGGAGGAGCTGAACCTGAGCTACAACGGCATCAC     2039
gi|126033153|gb|EF076731.1|       CAACACCTTCCTGGCTGTGCCCACCCTGGAGGAGCTGAACCTGAGCTACAACGGCATCAC     2039
gi|126033145|gb|EF076727.1|       CAACACCTTCCTGGCTGTGCCCACCCTGGAGGAGCTGAACCTGAGCTACAACGGCATCAC     2039
gi|126033141|gb|EF076725.1|       CAACACCTTCCTGGCTGTGCCCACCCTGGAGGAGCTGAACCTGAGCTACAACGGCATCAC     2039
gi|126033139|gb|EF076724.1|       CAACACCTTCCTGGCTGTGCCCACCCTGGAGGAGCTGAACCTGAGCTACAACGGCATCAC     2039
gi|126033149|gb|EF076729.1|       CAACACCTTCCTGGCTGTGCCCACCCTGGAGGAGCTGAACCTGAGCTACAACGGCATCAC     2039
gi|126033147|gb|EF076728.1|       CAACACCTTCCTGGCTGTGCCCACCCTGGAGGAGCTGAACCTGAGCTACAACGGCATCAC     2039
gi|126033137|gb|EF076723.1|       CAACACCTTCCTGGCTGTGCCCACCCTGGAGGAGCTGAACCTGAGCTACAACGGCATCAC     2039
gi|57471257|gb|AY859726.1|        CAACACCTTCCTGGCTGTGCCCACCCTGGAGGAGCTGAACCTGAGCTACAACGGCATCAC     487
gi|27848078|emb|AJ509824.2|       CAACACCTTCCTGGCTGTGCCCACCCTGGAGGAGCTGAACCTGAGCTACAACGGCATCAC     1767
gi|34147166|ref|NM_183081.1|      CAACACCTTCCTGGCTGTGCCCACCCTGGAGGAGCTGAACCTGAGCTACAACGGCATCAC     527
gi|23821159|emb|AJ509825.1|       CAACACCTTCCTGGCTGTGCCCACCCTGGAGGAGCTGAACCTGAGCTACAACGGCATCAC     383
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       GACCGTGCCTGCCCTGCCCAGTTCCCTCGTGTCCCTGTCGCTGAGCCACACCAGCATCCT     2084
gi|126033143|gb|EF076726.1|       GACCGTGCCTGCCCTGCCCAGTTCCCTCGTGTCCCTGTCGCTGAGCCACACCAGCATCCT     2099
gi|126033153|gb|EF076731.1|       GACCGTGCCTGCCCTGCCCAGTTCCCTCGTGTCCCTGTCGCTGAGCCACACCAGCATCCT     2099
gi|126033145|gb|EF076727.1|       GACCGTGCCTGCCCTGCCCAGTTCCCTCGTGTCCCTGTCGCTGAGCCACACCAGCATCCT     2099
gi|126033141|gb|EF076725.1|       GACCGTGCCTGCCCTGCCCAGTTCCCTCGTGTCCCTGTCGCTGAGCCACACCAGCATCCT     2099
gi|126033139|gb|EF076724.1|       GACCGTGCCTGCCCTGCCCAGTTCCCTCGTGTCCCTGTCGCTGAGCCACACCAGCATCCT     2099
gi|126033149|gb|EF076729.1|       GACCGTGCCTGCCCTGCCCAGTTCCCTCGTGTCCCTGTCGCTGAGCCACACCAGCATCCT     2099
gi|126033147|gb|EF076728.1|       GACCGTGCCTGCCCTGCCCAGTTCCCTCGTGTCCCTGTCGCTGAGCCACACCAGCATCCT     2099
gi|126033137|gb|EF076723.1|       GACCGTGCCTGCCCTGCCCAGTTCCCTCGTGTCCCTGTCGCTGAGCCACACCAGCATCCT     2099
gi|57471257|gb|AY859726.1|        GACCGTGCCTGCCCTGCCCAGTTCCCTCGTGTCCCTGTCGCTGAGCCACACCAGCATCCT     547
gi|27848078|emb|AJ509824.2|       GACCGTGCCTGCCCTGCCCAGTTCCCTCGTGTCCCTGTCGCTGAGCCACACCAGCATCCT     1827
gi|34147166|ref|NM_183081.1|      GACCGTGCCTGCCCTGCCCAGTTCCCTCGTGTCCCTGTCGCTGAGCCACACCAGCATCCT     587
gi|23821159|emb|AJ509825.1|       GACCGTGCCTGCCCTGCCCAGTTCCCTCGTGTCCCTGTCGCTGAGCCACACCAGCATCCT     443
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       GGTGCTAGGCCCCACCCACTTCACCGGCCTGCACGCCCTGCGCTTTCTGTACATGGACGG     2144
```

```
gi|126033143|gb|EF076726.1|       GGTGCTAGGCCCCACCCACTTCACCGGCCTGCACGCCCTGCGCTTTCTGTACATGGACGG     2159
gi|126033153|gb|EF076731.1|       GGTGCTAGGCCCCACCCACTTCACCGGCCTGCACGCCCTGCGCTTTCTGTACATGGACGG     2159
gi|126033145|gb|EF076727.1|       GGTGCTAGGCCCCACCCACTTCACCGGCCTGCACGCCCTGCGCTTTCTGTACATGGACGG     2159
gi|126033141|gb|EF076725.1|       GGTGCTAGGCCCCACCCACTTCACCGGCCTGCACGCCCTGCGCTTTCTGTACATGGACGG     2159
gi|126033139|gb|EF076724.1|       GGTGCTAGGCCCCACCCACTTCACCGGCCTGCACGCCCTGCGCTTTCTGTACATGGACGG     2159
gi|126033149|gb|EF076729.1|       GGTGCTAGGCCCCACCCACTTCACCGGCCTGCACGCCCTGCGCTTTCTGTACATGGACGG     2159
gi|126033147|gb|EF076728.1|       GGTGCTAGGCCCCACCCACTTCACCGGCCTGCACGCCCTGCGCTTTCTGTACATGGACGG     2159
gi|126033137|gb|EF076723.1|       GGTGCTAGGCCCCACCCACTTCACCGGCCTGCACGCCCTGCGCTTTCTGTACATGGACGG     2159
gi|57471257|gb|AY859726.1|        GGTGCTAGGCCCCACCCACTTCACCGGCCTGCACGCCCTGCGCTTTCTGTACATGGACGG     607
gi|27848078|emb|AJ509824.2|       GGTGCTAGGCCCCACCCACTTCACCGGCCTGCACGCCCTGCGCTTTCTGTACATGGACGG     1887
gi|34147166|ref|NM_183081.1|      GGTGCTAGGCCCCACCCACTTCACCGGCCTGCACGCCCTGCGCTTTCTGTACATGGACGG     647
gi|23821159|emb|AJ509825.1|       GGTGCTAGGCCCCACCCACTTCACCGGCCTGCACGCCCTGCGCTTTCTGTACATGGACGG     503
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       CAACTGCTACTACATGAACCCCTGCCCGCGGGCCCTGGAGGTGGCCCCAGGCGCCCTCCT     2204
gi|126033143|gb|EF076726.1|       CAACTGCTACTACATGAACCCCTGCCCGCGGGCCCTGGAGGTGGCCCCAGGCGCCCTCCT     2219
gi|126033153|gb|EF076731.1|       CAACTGCTACTACATGAACCCCTGCCCGCGGGCCCTGGAGGTGGCCCCAGGCGCCCTCCT     2219
gi|126033145|gb|EF076727.1|       CAACTGCTACTACATGAACCCCTGCCCGCGGGCCCTGGAGGTGGCCCCAGGCGCCCTCCT     2219
gi|126033141|gb|EF076725.1|       CAACTGCTACTACATGAACCCCTGCCCGCGGGCCCTGGAGGTGGCCCCAGGCGCCCTCCT     2219
gi|126033139|gb|EF076724.1|       CAACTGCTACTACATGAACCCCTGCCCGCGGGCCCTGGAGGTGGCCCCAGGCGCCCTCCT     2219
gi|126033149|gb|EF076729.1|       CAACTGCTACTACATGAACCCCTGCCCGCGGGCCCTGGAGGTGGCCCCAGGCGCCCTCCT     2219
gi|126033147|gb|EF076728.1|       CAACTGCTACTACATGAACCCCTGCCCGCGGGCCCTGGAGGTGGCCCCAGGCGCCCTCCT     2219
gi|126033137|gb|EF076723.1|       CAACTGCTACTACATGAACCCCTGCCCGCGGGCCCTGGAGGTGGCCCCAGGCGCCCTCCT     2219
gi|57471257|gb|AY859726.1|        CAACTGCTACTACATGAACCCCTGCCCGCGGGCCCTGGAGGTGGCCCCAGGCGCCCTCCT     667
gi|27848078|emb|AJ509824.2|       CAACTGCTACTACATGAACCCCTGCCCGCGGGCCCTGGAGGTGGCCCCAGGCGCCCTCCT     1947
gi|34147166|ref|NM_183081.1|      CAACTGCTACTACATGAACCCCTGCCCGCGGGCCCTGGAGGTGGCCCCAGGCGCCCTCCT     707
gi|23821159|emb|AJ509825.1|       CAACTGCTACTACATGAACCCCTGCCCGCGGGCCCTGGAGGTGGCCCCAGGCGCCCTCCT     563
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       CGGCCTGGGCAACCTCACGCACCTGTCGCTCAAGTACAACAACCTCACGGAGGTGCCCCG     2264
gi|126033143|gb|EF076726.1|       CGGCCTGGGCAACCTCACGCACCTGTCGCTCAAGTACAACAACCTCACGGAGGTGCCCCG     2279
gi|126033153|gb|EF076731.1|       CGGCCTGGGCAACCTCACGCACCTGTCGCTCAAGTACAACAACCTCACGGAGGTGCCCCG     2279
gi|126033145|gb|EF076727.1|       CGGCCTGGGCAACCTCACGCACCTGTCGCTCAAGTACAACAACCTCACGGAGGTGCCCCG     2279
gi|126033141|gb|EF076725.1|       CGGCCTGGGCAACCTCACGCACCTGTCGCTCAAGTACAACAACCTCACGGAGGTGCCCCG     2279
gi|126033139|gb|EF076724.1|       CGGCCTGGGCAACCTCACGCACCTGTCGCTCAAGTACAACAACCTCACGGAGGTGCCCCG     2279
gi|126033149|gb|EF076729.1|       CGGCCTGGGCAACCTCACGCACCTGTCGCTCAAGTACAACAACCTCACGGAGGTGCCCCG     2279
gi|126033147|gb|EF076728.1|       CGGCCTGGGCAACCTCACGCACCTGTCGCTCAAGTACAACAACCTCACGGAGGTGCCCCG     2279
gi|126033137|gb|EF076723.1|       CGGCCTGGGCAACCTCACGCACCTGTCGCTCAAGTACAACAACCTCACGGAGGTGCCCCG     2279
gi|57471257|gb|AY859726.1|        CGGCCTGGGCAACCTCACGCACCTGTCGCTCAAGTACAACAACCTCACGGAGGTGCCCCG     727
```

```
gi|27848078|emb|AJ509824.2|       CGGCCTGGGCAACCTCACGCACCTGTCGCTCAAGTACAACAACCTCACGGAGGTGCCCCG     2007
gi|34147166|ref|NM_183081.1|      CGGCCTGGGCAACCTCACGCACCTGTCGCTCAAGTACAACAACCTCACGGAGGTGCCCCG     767
gi|23821159|emb|AJ509825.1|       CGGCCTGGGCAACCTCACGCACCTGTCGCTCAAGTACAACAACCTCACGGAGGTGCCCCG     623
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       CCGCCTGCCCCCCAGCCTGGACACCCTGCTGCTGTCCTACAACCACATTGTCACCCTGGC     2324
gi|126033143|gb|EF076726.1|       CCGCCTGCCCCCCAGCCTGGACACCCTGCTGCTGTCCTACAACCACATTGTCACCCTGGC     2339
gi|126033153|gb|EF076731.1|       CCGCCTGCCCCCCAGCCTGGACACCCTGCTGCTGTCCTACAACCACATTGTCACCCTGGC     2339
gi|126033145|gb|EF076727.1|       CCGCCTGCCCCCCAGCCTGGACACCCTGCTGCTGTCCTACAACCACATTGTCACCCTGGC     2339
gi|126033141|gb|EF076725.1|       CCGCCTGCCCCCCAGCCTGGACACCCTGCTGCTGTCCTACAACCACATTGTCACCCTGGC     2339
gi|126033139|gb|EF076724.1|       CCGCCTGCCCCCCAGCCTGGACACCCTGCTGCTGTCCTACAACCACATTGTCACCCTGGC     2339
gi|126033149|gb|EF076729.1|       CCGCCTGCCCCCCAGCCTGGACACCCTGCTGCTGTCCTACAACCACATTGTCACCCTGGC     2339
gi|126033147|gb|EF076728.1|       CCGCCTGCCCCCCAGCCTGGACACCCTGCTGCTGTCCTACAACCACATTGTCACCCTGGC     2339
gi|126033137|gb|EF076723.1|       CCGCCTGCCCCCCAGCCTGGACACCCTGCTGCTGTCCTACAACCACATTGTCACCCTGGC     2339
gi|57471257|gb|AY859726.1|        CCGCCTGCCCCCCAGCCTGGACACCCTGCTGCTGTCCTACAACCACATTGTCACCCTGGC     787
gi|27848078|emb|AJ509824.2|       CCGCCTGCCCCCCAGCCTGGACACCCTGCTGCTGTCCTACAACCACATTGTCACCCTGGC     2067
gi|34147166|ref|NM_183081.1|      CCGCCTGCCCCCCAGCCTGGACACCCTGCTGCTGTCCTACAACCACATTGTCACCCTGGC     827
gi|23821159|emb|AJ509825.1|       CCGCCTGCCCCCCAGCCTGGACACCCTGCTGCTGTCCTACAACCACATTGTCACCCTGGC     683
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       ACCCGAGGACCTGGCCAACCTGACTGCCCTGCGCGTGCTTGACGTGGGTGGGAACTGCCG     2384
gi|126033143|gb|EF076726.1|       ACCCGAGGACCTGGCCAACCTGACTGCCCTGCGCGTGCTTGACGTGGGTGGGAACTGCCG     2399
gi|126033153|gb|EF076731.1|       ACCCGAGGACCTGGCCAACCTGACTGCCCTGCGCGTGCTTGACGTGGGTGGGAACTGCCG     2399
gi|126033145|gb|EF076727.1|       ACCCGAGGACCTGGCCAACCTGACTGCCCTGCGCGTGCTTGACGTGGGTGGGAACTGCCG     2399
gi|126033141|gb|EF076725.1|       ACCCGAGGACCTGGCCAACCTGACTGCCCTGCGCGTGCTTGACGTGGGTGGGAACTGCCG     2399
gi|126033139|gb|EF076724.1|       ACCCGAGGACCTGGCCAACCTGACTGCCCTGCGCGTGCTTGACGTGGGTGGGAACTGCCG     2399
gi|126033149|gb|EF076729.1|       ACCCGAGGACCTGGCCAACCTGACTGCCCTGCGCGTGCTTGACGTGGGTGGGAACTGCCG     2399
gi|126033147|gb|EF076728.1|       ACCCGAGGACCTGGCCAACCTGACTGCCCTGCGCGTGCTTGACGTGGGTGGGAACTGCCG     2399
gi|126033137|gb|EF076723.1|       ACCCGAGGACCTGGCCAACCTGACTGCCCTGCGCGTGCTTGACGTGGGTGGGAACTGCCG     2399
gi|57471257|gb|AY859726.1|        ACCCGAGGACCTGGCCAACCTGACTGCCCTGCGCGTGCTTGACGTGGGTGGGAACTGCCG     847
gi|27848078|emb|AJ509824.2|       ACCCGAGGACCTGGCCAACCTGACTGCCCTGCGCGTGCTCGACGTGGGTGGGAACTGCCG     2127
gi|34147166|ref|NM_183081.1|      ACCCGAGGACCTGGCCAACCTGACTGCCCTGCGCGTGCTCGACGTGGGTGGGAACTGCCG     887
gi|23821159|emb|AJ509825.1|       ACCCGAGGACCTGGCCAACCTGACTGCCCTGCGCGTGCTCGACGTGGGTGGGAACTGCCG     743
                                  ************************************ ********************

gi|407316560|gb|FJ495080.2|       CCGCTGCGACCATGCCCGCAACCCCTGCAGGGAGTGCCCAAAGAACTTCCCCAAGCTGCA     2444
gi|126033143|gb|EF076726.1|       CCGCTGCGACCATGCCCGCAACCCCTGCAGGGAGTGCCCAAAGAACTTCCCCAAGCTGCA     2459
gi|126033153|gb|EF076731.1|       CCGCTGCGACCATGCCCGCAACCCCTGCAGGGAGTGCCCAAAGAACTTCCCCAAGCTGCA     2459
gi|126033145|gb|EF076727.1|       CCGCTGCGACCATGCCCGCAACCCCTGCAGGGAGTGCCCAAAGAACTTCCCCAAGCTGCA     2459
```

```
gi|126033141|gb|EF076725.1|       CCGCTGCGACCATGCCCGCAACCCCTGCAGGGAGTGCCCAAAGAACTTCCCCAAGCTGCA     2459
gi|126033139|gb|EF076724.1|       CCGCTGCGACCATGCCCGCAACCCCTGCAGGGAGTGCCCAAAGAACTTCCCCAAGCTGCA     2459
gi|126033149|gb|EF076729.1|       CCGCTGCGACCATGCCCGCAACCCCTGCAGGGAGTGCCCAAAGAACTTCCCCAAGCTGCA     2459
gi|126033147|gb|EF076728.1|       CCGCTGCGACCATGCCCGCAACCCCTGCAGGGAGTGCCCAAAGAACTTCCCCAAGCTGCA     2459
gi|126033137|gb|EF076723.1|       CCGCTGCGACCATGCCCGCAACCCCTGCAGGGAGTGCCCAAAGAACTTCCCCAAGCTGCA     2459
gi|57471257|gb|AY859726.1|        CCGCTGCGACCATGCCCGCAACCCCTGCAGGGAGTGCCCAAAGAACTTCCCCAAGCTGCA     907
gi|27848078|emb|AJ509824.2|       CCGCTGCGACCATGCCCGCAACCCCTGCAGGGAGTGCCCAAAGAACTTCCCCAAGCTGCA     2187
gi|34147166|ref|NM_183081.1|      CCGCTGCGACCATGCCCGCAACCCCTGCAGGGAGTGCCCAAAGAACTTCCCCAAGCTGCA     947
gi|23821159|emb|AJ509825.1|       CCGCTGCGACCATGCCCGCAACCCCTGCAGGGAGTGCCCAAAGAACTTCCCCAAGCTGCA     803
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       CCCTGACACCTTCAGTCACCTGAGCCGCCTCGAAGGCCTGGTGTTGAAGGACAGTTCTCT     2504
gi|126033143|gb|EF076726.1|       CCCTGACACCTTCAGTCACCTGAGCCGCCTCGAAGGCCTGGTGTTGAAGGACAGTTCTCT     2519
gi|126033153|gb|EF076731.1|       CCCTGACACCTTCAGTCACCTGAGCCGCCTCGAAGGCCTGGTGTTGAAGGACAGTTCTCT     2519
gi|126033145|gb|EF076727.1|       CCCTGACACCTTCAGTCACCTGAGCCGCCTCGAAGGCCTGGTGTTGAAGGACAGTTCTCT     2519
gi|126033141|gb|EF076725.1|       CCCTGACACCTTCAGTCACCTGAGCCGCCTCGAAGGCCTGGTGTTGAAGGACAGTTCTCT     2519
gi|126033139|gb|EF076724.1|       CCCTGACACCTTCAGTCACCTGAGCCGCCTCGAAGGCCTGGTGTTGAAGGACAGTTCTCT     2519
gi|126033149|gb|EF076729.1|       CCCTGACACCTTCAGTCACCTGAGCCGCCTCGAAGGCCTGGTGTTGAAGGACAGTTCTCT     2519
gi|126033147|gb|EF076728.1|       CCCTGACACCTTCAGTCACCTGAGCCGCCTCGAAGGCCTGGTGTTGAAGGACAGTTCTCT     2519
gi|126033137|gb|EF076723.1|       CCCTGACACCTTCAGTCACCTGAGCCGCCTCGAAGGCCTGGTGTTGAAGGACAGTTCTCT     2519
gi|57471257|gb|AY859726.1|        CCCTGACACCTTCAGTCACCTGAGCCGCCTCGAAGGCCTGGTGTTGAAGGACAGTTCTCT     967
gi|27848078|emb|AJ509824.2|       CCCTGACACCTTCAGTCACCTGAGCCGCCTCGAAGGCCTGGTGTTGAAGGACAGTTCTCT     2247
gi|34147166|ref|NM_183081.1|      CCCTGACACCTTCAGTCACCTGAGCCGCCTCGAAGGCCTGGTGTTGAAGGACAGTTCTCT     1007
gi|23821159|emb|AJ509825.1|       CCCTGACACCTTCAGTCACCTGAGCCGCCTCGAAGGCCTGGTGTTGAAGGACAGTTCTCT     863
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       CTACAAACTAGAGAAAGATTGGTTCCGCGGCCTGGGCAGGCTCCAAGTGCTCGACCTGAG     2564
gi|126033143|gb|EF076726.1|       CTACAAACTAGAGAAAGATTGGTTCCGCGGCCTGGGCAGGCTCCAAGTGCTCGACCTGAG     2579
gi|126033153|gb|EF076731.1|       CTACAAACTAGAGAAAGATTGGTTCCGCGGCCTGGGCAGGCTCCAAGTGCTCGACCTGAG     2579
gi|126033145|gb|EF076727.1|       CTACAAACTAGAGAAAGATTGGTTCCGCGGCCTGGGCAGGCTCCAAGTGCTCGACCTGAG     2579
gi|126033141|gb|EF076725.1|       CTACAAACTAGAGAAAGATTGGTTCCGCGGCCTGGGCAGGCTCCAAGTGCTCGACCTGAG     2579
gi|126033139|gb|EF076724.1|       CTACAAACTAGAGAAAGATTGGTTCCGCGGCCTGGGCAGGCTCCAAGTGCTCGACCTGAG     2579
gi|126033149|gb|EF076729.1|       CTACAAACTAGAGAAAGATTGGTTCCGCGGCCTGGGCAGGCTCCAAGTGCTCGACCTGAG     2579
gi|126033147|gb|EF076728.1|       CTACAAACTAGAGAAAGATTGGTTCCGCGGCCTGGGCAGGCTCCAAGTGCTCGACCTGAG     2579
gi|126033137|gb|EF076723.1|       CTACAAACTAGAGAAAGATTGGTTCCGCGGCCTGGGCAGGCTCCAAGTGCTCGACCTGAG     2579
gi|57471257|gb|AY859726.1|        CTACAAACTAGAGAAAGATTGGTTCCGCGGCCTGGGCAGGCTCCAAGTGCTCGACCTGAG     1027
gi|27848078|emb|AJ509824.2|       CTACAAACTAGAGAAAGATTGGTTCCGCGGCCTGGGCAGGCTCCAAGTGCTCGACCTGAG     2307
gi|34147166|ref|NM_183081.1|      CTACAAACTAGAGAAAGATTGGTTCCGCGGCCTGGGCAGGCTCCAAGTGCTCGACCTGAG     1067
gi|23821159|emb|AJ509825.1|       CTACAAACTAGAGAAAGATTGGTTCCGCGGCCTGGGCAGGCTCCAAGTGCTCGACCTGAG     923
```

```
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       TGAGAACTTCCTCTATGACTACATCACCAAGACCACCATCTTCAACGACCTGACCCAGCT     2624
gi|126033143|gb|EF076726.1|       TGAGAACTTCCTCTATGACTACATCACCAAGACCACCATCTTCAACGACCTGACCCAGCT     2639
gi|126033153|gb|EF076731.1|       TGAGAACTTCCTCTATGACTACATCACCAAGACCACCATCTTCAACGACCTGACCCAGCT     2639
gi|126033145|gb|EF076727.1|       TGAGAACTTCCTCTATGACTACATCACCAAGACCACCATCTTCAACGACCTGACCCAGCT     2639
gi|126033141|gb|EF076725.1|       TGAGAACTTCCTCTATGACTACATCACCAAGACCACCATCTTCAACGACCTGACCCAGCT     2639
gi|126033139|gb|EF076724.1|       TGAGAACTTCCTCTATGACTACATCACCAAGACCACCATCTTCAACGACCTGACCCAGCT     2639
gi|126033149|gb|EF076729.1|       TGAGAACTTCCTCTATGACTACATCACCAAGACCACCATCTTCAACGACCTGACCCAGCT     2639
gi|126033147|gb|EF076728.1|       TGAGAACTTCCTCTATGACTACATCACCAAGACCACCATCTTCAACGACCTGACCCAGCT     2639
gi|126033137|gb|EF076723.1|       TGAGAACTTCCTCTATGACTACATCACCAAGACCACCATCTTCAACGACCTGACCCAGCT     2639
gi|57471257|gb|AY859726.1|        TGAGAACTTCCTCTATGACTACATCACCAAGACCACCATCTTCAACGACCTGACCCAGCT     1087
gi|27848078|emb|AJ509824.2|       TGAGAACTTCCTCTATGACTACATCACCAAGACCACCATCTTCAACGACCTGACCCAGCT     2367
gi|34147166|ref|NM_183081.1|      TGAGAACTTCCTCTATGACTACATCACCAAGACCACCATCTTCAACGACCTGACCCAGCT     1127
gi|23821159|emb|AJ509825.1|       TGAGAACTTCCTCTATGACTACATCACCAAGACCACCATCTTCAACGACCTGACCCAGCT     983
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       GCGCAGACTCAACCTGTCCTTCAATTACCACAAGAAGGTGTCCTTCGCCCACCTGCACCT     2684
gi|126033143|gb|EF076726.1|       GCGCAGACTCAACCTGTCCTTCAATTACCACAAGAAGGTGTCCTTCGCCCACCTGCACCT     2699
gi|126033153|gb|EF076731.1|       GCGCAGACTCAACCTGTCCTTCAATTACCACAAGAAGGTGTCCTTCGCCCACCTGCACCT     2699
gi|126033145|gb|EF076727.1|       GCGCAGACTCAACCTGTCCTTCAATTACCACAAGAAGGTGTCCTTCGCCCACCTGCACCT     2699
gi|126033141|gb|EF076725.1|       GCGCAGACTCAACCTGTCCTTCAATTACCACAAGAAGGTGTCCTTCGCCCACCTGCACCT     2699
gi|126033139|gb|EF076724.1|       GCGCAGACTCAACCTGTCCTTCAATTACCACAAGAAGGTGTCCTTCGCCCACCTGCACCT     2699
gi|126033149|gb|EF076729.1|       GCGCAGACTCAACCTGTCCTTCAATTACCACAAGAAGGTGTCCTTCGCCCACCTGCACCT     2699
gi|126033147|gb|EF076728.1|       GCGCAGACTCAACCTGTCCTTCAATTACCACAAGAAGGTGTCCTTCGCCCACCTGCACCT     2699
gi|126033137|gb|EF076723.1|       GCGCAGACTCAACCTGTCCTTCAATTACCACAAGAAGGTGTCCTTCGCCCACCTGCACCT     2699
gi|57471257|gb|AY859726.1|        GCGCAGACTCAACCTGTCCTTCAATTACCACAAGAAGGTGTCCTTCGCCCACCTGCACCT     1147
gi|27848078|emb|AJ509824.2|       GCGCAGACTCAACCTGTCCTTCAATTACCACAAGAAGGTGTCCTTCGCCCACCTGCACCT     2427
gi|34147166|ref|NM_183081.1|      GCGCAGACTCAACCTGTCCTTCAATTACCACAAGAAGGTGTCCTTCGCCCACCTGCACCT     1187
gi|23821159|emb|AJ509825.1|       GCGCAGACTCAACCTGTCCTTCAATTACCACAAGAAGGTGTCCTTCGCCCACCTGCACCT     1043
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       GGCGTCCTCCTTTGGGAGTCTGGTGTCCCTGGAGAAGCTGGACATGCACGGCATCTTCTT     2744
gi|126033143|gb|EF076726.1|       GGCGTCCTCCTTTGGGAGTCTGGTGTCCCTGGAGAAGCTGGACATGCACGGCATCTTCTT     2759
gi|126033153|gb|EF076731.1|       GGCGTCCTCCTTTGGGAGTCTGGTGTCCCTGGAGAAGCTGGACATGCACGGCATCTTCTT     2759
gi|126033145|gb|EF076727.1|       GGCGTCCTCCTTTGGGAGTCTGGTGTCCCTGGAGAAGCTGGACATGCACGGCATCTTCTT     2759
gi|126033141|gb|EF076725.1|       GGCGTCCTCCTTTGGGAGTCTGGTGTCCCTGGAGAAGCTGGACATGCACGGCATCTTCTT     2759
gi|126033139|gb|EF076724.1|       GGCGTCCTCCTTTGGGAGTCTGGTGTCCCTGGAGAAGCTGGACATGCACGGCATCTTCTT     2759
gi|126033149|gb|EF076729.1|       GGCGTCCTCCTTTGGGAGTCTGGTGTCCCTGGAGAAGCTGGACATGCACGGCATCTTCTT     2759
```

```
gi|126033147|gb|EF076728.1|       GGCGTCCTCCTTTGGGAGTCTGGTGTCCCTGGAGAAGCTGGACATGCACGGCATCTTCTT     2759
gi|126033137|gb|EF076723.1|       GGCGTCCTCCTTTGGGAGTCTGGTGTCCCTGGAGAAGCTGGACATGCACGGCATCTTCTT     2759
gi|57471257|gb|AY859726.1|        AGCGTCCTCCTTTGGGAGTCTGGTGTCCCTGGAGAAGCTGGACATGCACGGCATCTTCTT     1207
gi|27848078|emb|AJ509824.2|       GGCGTCCTCCTTTGGGAGTCTGGTGTCCCTGGAGAAGCTGGACATGCACGGCATCTTCTT     2487
gi|34147166|ref|NM_183081.1|      GGCGTCCTCCTTTGGGAGTCTGGTGTCCCTGGAGAAGCTGGACATGCACGGCATCTTCTT     1247
gi|23821159|emb|AJ509825.1|       GGCGTCCTCCTTTGGGAGTCTGGTGTCCCTGGAGAAGCTGGACATGCACGGCATCTTCTT     1103
                                   ************************************************************

gi|407316560|gb|FJ495080.2|       CCGCTCCCTCACCAACATCACGCTCCAGCCGCTGACCCGGCTGCCCAAGCTCCAGAGTCT     2804
gi|126033143|gb|EF076726.1|       CCGCTCCCTCACCAACATCACGCTCCAGCCGCTGACCCGGCTGCCCAAGCTCCAGAGTCT     2819
gi|126033153|gb|EF076731.1|       CCGCTCCCTCACCAACATCACGCTCCAGCCGCTGACCCGGCTGCCCAAGCTCCAGAGTCT     2819
gi|126033145|gb|EF076727.1|       CCGCTCCCTCACCAACATCACGCTCCAGCCGCTGACCCGGCTGCCCAAGCTCCAGAGTCT     2819
gi|126033141|gb|EF076725.1|       CCGCTCCCTCACCAACATCACGCTCCAGCCGCTGACCCGGCTGCCCAAGCTCCAGAGTCT     2819
gi|126033139|gb|EF076724.1|       CCGCTCCCTCACCAACATCACGCTCCAGCCGCTGACCCGGCTGCCCAAGCTCCAGAGTCT     2819
gi|126033149|gb|EF076729.1|       CCGCTCCCTCACCAACATCACGCTCCAGCCGCTGACCCGGCTGCCCAAGCTCCAGAGTCT     2819
gi|126033147|gb|EF076728.1|       CCGCTCCCTCACCAACATCACGCTCCAGCCGCTGACCCGGCTGCCCAAGCTCCAGAGTCT     2819
gi|126033137|gb|EF076723.1|       CCGCTCCCTCACCAACATCACGCTCCAGCCGCTGACCCGGCTGCCCAAGCTCCAGAGTCT     2819
gi|57471257|gb|AY859726.1|        CCGCTCCCTCACCAACATCACGCTCCAGTCGCTGACCCGGCTGCCCAAGCTCCAGAGTCT     1267
gi|27848078|emb|AJ509824.2|       CCGCTCCCTCACCAACATCACGCTCCAGCCGCTGACCCGGCTGCCCAAGCTCCAGAGTCT     2547
gi|34147166|ref|NM_183081.1|      CCGCTCCCTCACCAACATCACGCTCCAGCCGCTGACCCGGCTGCCCAAGCTCCAGAGTCT     1307
gi|23821159|emb|AJ509825.1|       CCGCTCCCTCACCAACATCACGCTCCAGCCGCTGACCCGGCTGCCCAAGCTCCAGAGTCT     1163
                                   *************************** ******************************

gi|407316560|gb|FJ495080.2|       GCGTCTGCAGCTGAACTTCATCAACCAGGCCCAGCTCAGCATCTTTGGGGCCTTCCCGAG     2864
gi|126033143|gb|EF076726.1|       GCGTCTGCAGCTGAACTTCATCAACCAGGCCCAGCTCAGCATCTTTGGGGCCTTCCCGAG     2879
gi|126033153|gb|EF076731.1|       GCGTCTGCAGCTGAACTTCATCAACCAGGCCCAGCTCAGCATCTTTGGGGCCTTCCCGAG     2879
gi|126033145|gb|EF076727.1|       GCGTCTGCAGCTGAACTTCATCAACCAGGCCCAGCTCAGCATCTTTGGGGCCTTCCCGAG     2879
gi|126033141|gb|EF076725.1|       GCGTCTGCAGCTGAACTTCATCAACCAGGCCCAGCTCAGCATCTTTGGGGCCTTCCCGAG     2879
gi|126033139|gb|EF076724.1|       GCGTCTGCAGCTGAACTTCATCAACCAGGCCCAGCTCAGCATCTTTGGGGCCTTCCCGAG     2879
gi|126033149|gb|EF076729.1|       GCGTCTGCAGCTGAACTTCATCAACCAGGCCCAGCTCAGCATCTTTGGGGCCTTCCCGAG     2879
gi|126033147|gb|EF076728.1|       GCGTCTGCAGCTGAACTTCATCAACCAGGCCCAGCTCAGCATCTTTGGGGCCTTCCCGAG     2879
gi|126033137|gb|EF076723.1|       GCGTCTGCAGCTGAACTTCATCAACCAGGCCCAGCTCAGCATCTTTGGGGCCTTCCCGAG     2879
gi|57471257|gb|AY859726.1|        GCATCTGCAGCTGAACTTCATCAACCAGGCCCAGCTCAGCATCTTTGGGGCCTTCCCGAG     1327
gi|27848078|emb|AJ509824.2|       GCGTCTGCAGCTGAACTTCATCAACCAGGCCCAGCTCAGCATCTTTGGGGCCTTCCCGAG     2607
gi|34147166|ref|NM_183081.1|      GCGTCTGCAGCTGAACTTCATCAACCAGGCCCAGCTCAGCATCTTTGGGGCCTTCCCGAG     1367
gi|23821159|emb|AJ509825.1|       GCGTCTGCAGCTGAACTTCATCAACCAGGCCCAGCTCAGCATCTTTGGGGCCTTCCCGAG     1223
                                   ** ********************************************************

gi|407316560|gb|FJ495080.2|       CCTGCTCTTCGTGGACCTGTCGGACAACCGCATCAGCGGAGCCGCGACGCCGGCGGCCGC     2924
```

```
gi|126033143|gb|EF076726.1|       CCTGCTCTTCGTGGACCTGTCGGACAACCGCATCAGCGGAGCCGCGACGCCAGCGGCCGC     2939
gi|126033153|gb|EF076731.1|       CCTGCTCTTCGTGGACCTGTCGGACAACCGCATCAGCGGAGCCGCGACGCCAGCGGCCGC     2939
gi|126033145|gb|EF076727.1|       CCTGCTCTTCGTGGACCTGTCGGACAACCGCATCAGCGGAGCCGCGACGCCAGCGGCCGC     2939
gi|126033141|gb|EF076725.1|       CCTGCTCTTCGTGGACCTGTCGGACAACCGCATCAGCGGAGCCGCGACGCCAGCGGCCGC     2939
gi|126033139|gb|EF076724.1|       CCTGCTCTTCGTGGACCTGTCGGACAACCGCATCAGCGGAGCCGCGACGCCAGCGGCCGC     2939
gi|126033149|gb|EF076729.1|       CCTGCTCTTCGTGGACCTGTCGGACAACCGCATCAGCGGAGCCGCGACGCCAGCGGCCGC     2939
gi|126033147|gb|EF076728.1|       CCTGCTCTTCGTGGACCTGTCGGACAACCGCATCAGCGGAGCCGCGACGCCAGCGGCCGC     2939
gi|126033137|gb|EF076723.1|       CCTGCTCTTCGTGGACCTGTCGGACAACCGCATCAGCGGAGCCGCGACGCCAGCGGCCGC     2939
gi|57471257|gb|AY859726.1|        CCTGCTCTTCGTGGACCTGTCGGACAACCGCATCAGCGGAGCCGCGACGCCAGCGGCCGC     1387
gi|27848078|emb|AJ509824.2|       CCTGCTCTTCGTGGACCTGTCGGACAACCGCATCAGCGGAGCCGCGACGCCAGCGGCCGC     2667
gi|34147166|ref|NM_183081.1|      CCTGCTCTTCGTGGACCTGTCGGACAACCGCATCAGCGGAGCCGCGACGCCAGCGGCCGC     1427
gi|23821159|emb|AJ509825.1|       CCTGCTCTTCGTGGACCTGTCGGACAACCGCATCAGCGGAGCCGCGACGCCAGCGGCCGC     1283
                                  ************************************************ ********

gi|407316560|gb|FJ495080.2|       CCTGGAGGAGGTGGACAGCAGGGTGGAAGTCTGGCGATTGCCCAGGGGCCTCGCTCCAGG     2984
gi|126033143|gb|EF076726.1|       CCTGGGGGAGGTGGACAGCAGGGTGGAAGTCTGGCGATTGCCCAGGGGCCTCGCTCCAGG     2999
gi|126033153|gb|EF076731.1|       CCTGGGGGAGGTGGACAGCAGGGTGGAAGTCTGGCGATTGCCCAGGGGCCTCGCTCCAGG     2999
gi|126033145|gb|EF076727.1|       CCTGGGGGAGGTGGACAGCAGGGTGGAAGTCTGGCGATTGCCCAGGGGCCTCGCTCCAGG     2999
gi|126033141|gb|EF076725.1|       CCTGGGGGAGGTGGACAGCAGGGTGGAAGTCTGGCGATTGCCCAGGGGCCTCGCTCCAGG     2999
gi|126033139|gb|EF076724.1|       CCTGGGGGAGGTGGACAGCAGGGTGGAAGTCTGGCGATTGCCCAGGGGCCTCGCTCCAGG     2999
gi|126033149|gb|EF076729.1|       CCTGGGGGAGGTGGACAGCAGGGTGGAAGTCTGGCGATTGCCCAGGGGCCTCGCTCCAGG     2999
gi|126033147|gb|EF076728.1|       CCTGGGGGAGGTGGACAGCAGGGTGGAAGTCTGGCGATTGCCCAGGGGCCTCGCTCCAGG     2999
gi|126033137|gb|EF076723.1|       CCTGGGGGAGGTGGACAGCAGGGTGGAAGTCTGGCGATTGCCCAGGGGCCTCGCTCCAGG     2999
gi|57471257|gb|AY859726.1|        CCTGGGGGAGGTGGACAGCAGGGTGGAAGTCTGGCGATTGCCCAGGGGCCTCGCTCCAGG     1447
gi|27848078|emb|AJ509824.2|       CCTGGGGGAGGTGGACAGCAGGGTGGAAGTCTGGCGATTGCCCAGGGGCCTCGCTCCAGG     2727
gi|34147166|ref|NM_183081.1|      CCTGGGGGAGGTGGACAGCAGGGTGGAAGTCTGGCGATTGCCCAGGGGCCTCGCTCCAGG     1487
gi|23821159|emb|AJ509825.1|       CCTGGGGGAGGTGGACAGCAGGGTGGAAGTCTGGCGATTGCCCAGGGGCCTCGCTCCAGG     1343
                                  ***** **************************************************

gi|407316560|gb|FJ495080.2|       CCCGCTGGACGCCGTCAGCTCAAAGGACTTCATGCCAAGCTGCAACCTCAACTTCACCTT     3044
gi|126033143|gb|EF076726.1|       CCCGCTGGACGCCGTCAGCTCAAAGGACTTCATGCCAAGCTGCAACCTCAACTTCACCTT     3059
gi|126033153|gb|EF076731.1|       CCCGCTGGACGCCGTCAGCTCAAAGGACTTCATGCCAAGCTGCAACCTCAACTTCACCTT     3059
gi|126033145|gb|EF076727.1|       CCCGCTGGACGCCGTCAGCTCAAAGGACTTCATGCCAAGCTGCAACCTCAACTTCACCTT     3059
gi|126033141|gb|EF076725.1|       CCCGCTGGACGCCGTCAGCTCAAAGGACTTCATGCCAAGCTGCAACCTCAACTTCACCTT     3059
gi|126033139|gb|EF076724.1|       CCCGCTGGACGCCGTCAGCTCAAAGGACTTCATGCCAAGCTGCAACCTCAACTTCACCTT     3059
gi|126033149|gb|EF076729.1|       CCCGCTGGACGCCGTCAGCTCAAAGGACTTCATGCCAAGCTGCAACCTCAACTTCACCTT     3059
gi|126033147|gb|EF076728.1|       CCCGCTGGACGCCGTCAGCTCAAAGGACTTCATGCCAAGCTGCAACCTCAACTTCACCTT     3059
gi|126033137|gb|EF076723.1|       CCCGCTGGACGCCGTCAGCTCAAAGGACTTCATGCCAAGCTGCAACCTCAACTTCACCTT     3059
gi|57471257|gb|AY859726.1|        CCCGCTGGACGCCGTCAGCTCAAAGGACTTCATGCCAAGCTGCAACCTCAACTTCACCTT     1507
```

```
gi|27848078|emb|AJ509824.2|       CCCGCTGGACGCCGTCAGCTCAAAGGACTTCATGCCAAGCTGCAACCTCAACTTCACCTT     2787
gi|34147166|ref|NM_183081.1|      CCCGCTGGACGCCGTCAGCTCAAAGGACTTCATGCCAAGCTGCAACCTCAACTTCACCTT     1547
gi|23821159|emb|AJ509825.1|       CCCGCTGGACGCCGTCAGCTCAAAGGACTTCATGCCAAGCTGCAACCTCAACTTCACCTT     1403
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       GGACCTGTCACGGAACAACCTGGTGACAATCCAGCAAGAGATGTTTACCCGCCTCTCCCG     3104
gi|126033143|gb|EF076726.1|       GGACCTGTCACGGAACAACCTGGTGACAATCCAGCAAGAGATGTTTACCCGCCTCTCCCG     3119
gi|126033153|gb|EF076731.1|       GGACCTGTCACGGAACAACCTGGTGACAATCCAGCAAGAGATGTTTACCCGCCTCTCCCG     3119
gi|126033145|gb|EF076727.1|       GGACCTGTCACGGAACAACCTGGTGACAATCCAGCAAGAGATGTTTACCCGCCTCTCCCG     3119
gi|126033141|gb|EF076725.1|       GGACCTGTCACGGAACAACCTGGTGACAATCCAGCAAGAGATGTTTACCCGCCTCTCCCG     3119
gi|126033139|gb|EF076724.1|       GGACCTGTCACGGAACAACCTGGTGACAATCCAGCAAGAGATGTTTACCCGCCTCTCCCG     3119
gi|126033149|gb|EF076729.1|       GGACCTGTCACGGAACAACCTGGTGACAATCCAGCAAGAGATGTTTACCCGCCTCTCCCG     3119
gi|126033147|gb|EF076728.1|       GGACCTGTCACGGAACAACCTGGTGACAATCCAGCAAGAGATGTTTACCCGCCTCTCCCG     3119
gi|126033137|gb|EF076723.1|       GGACCTGTCACGGAACAACCTGGTGACAATCCAGCAAGAGATGTTTACCCGCCTCTCCCG     3119
gi|57471257|gb|AY859726.1|        GGACCTGTCACGGAACAACCTGGTGACAATCCAGCAAGAGATGTTTACCCGCCTCTCCCG     1567
gi|27848078|emb|AJ509824.2|       GGACCTGTCACGGAACAACCTGGTGACAATCCAGCAAGAGATGTTTACCCGCCTCTCCCG     2847
gi|34147166|ref|NM_183081.1|      GGACCTGTCACGGAACAACCTGGTGACAATCCAGCAAGAGATGTTTACCCGCCTCTCCCG     1607
gi|23821159|emb|AJ509825.1|       GGACCTGTCACGGAACAACCTGGTGACAATCCAGCAAGAGATGTTTACCCGCCTCTCCCG     1463
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       CCTCCAGTGCCTGCGCCTGAGCCACAACAGCATCTCACAGGCGGTTAATGGCTCCCAGTT     3164
gi|126033143|gb|EF076726.1|       CCTCCAGTGCCTGCGCCTGAGCCACAACAGCATCTCACAGGCGGTTAATGGCTCCCAGTT     3179
gi|126033153|gb|EF076731.1|       CCTCCAGTGCCTGCGCCTGAGCCACAACAGCATCTCACAGGCGGTTAATGGCTCCCAGTT     3179
gi|126033145|gb|EF076727.1|       CCTCCAGTGCCTGCGCCTGAGCCACAACAGCATCTCACAGGCGGTTAATGGCTCCCAGTT     3179
gi|126033141|gb|EF076725.1|       CCTCCAGTGCCTGCGCCTGAGCCACAACAGCATCTCACAGGCGGTTAATGGCTCCCAGTT     3179
gi|126033139|gb|EF076724.1|       CCTCCAGTGCCTGCGCCTGAGCCACAACAGCATCTCACAGGCGGTTAATGGCTCCCAGTT     3179
gi|126033149|gb|EF076729.1|       CCTCCAGTGCCTGCGCCTGAGCCACAACAGCATCTCACAGGCGGTTAATGGCTCCCAGTT     3179
gi|126033147|gb|EF076728.1|       CCTCCAGTGCCTGCGCCTGAGCCACAACAGCATCTCACAGGCGGTTAATGGCTCCCAGTT     3179
gi|126033137|gb|EF076723.1|       CCTCCAGTGCCTGCGCCTGAGCCACAACAGCATCTCACAGGCGGTTAATGGCTCCCAGTT     3179
gi|57471257|gb|AY859726.1|        CCTCCAGTGCCTGCGCCTGAGCCACAACAGCATCTCGCAGGCGGTTAATGGCTCCCAGTT     1627
gi|27848078|emb|AJ509824.2|       CCTCCAGTGCCTGCGCCTGAGCCACAACAGCATCTCACAGGCGGTTAATGGCTCCCAGTT     2907
gi|34147166|ref|NM_183081.1|      CCTCCAGTGCCTGCGCCTGAGCCACAACAGCATCTCACAGGCGGTTAATGGCTCCCAGTT     1667
gi|23821159|emb|AJ509825.1|       CCTCCAGTGCCTGCGCCTGAGCCACAACAGCATCTCACAGGCGGTTAATGGCTCCCAGTT     1523
                                  ********************************** **********************

gi|407316560|gb|FJ495080.2|       CGTGCCGCTGACCAGCCTGCGAGTGCTCGACCTGTCCCACAACAAGCTGGACCTGTACCA     3224
gi|126033143|gb|EF076726.1|       CGTGCCGCTGACCAGCCTGCGAGTGCTCGACCTGTCCCACAACAAGCTGGACCTGTACCA     3239
gi|126033153|gb|EF076731.1|       CGTGCCGCTGACCAGCCTGCGAGTGCTCGACCTGTCCCACAACAAGCTGGACCTGTACCA     3239
gi|126033145|gb|EF076727.1|       CGTGCCGCTGACCAGCCTGCGAGTGCTCGACCTGTCCCACAACAAGCTGGACCTGTACCA     3239
```

```
gi|126033141|gb|EF076725.1|       CGTGCCGCTGACCAGCCTGCGAGTGCTCGACCTGTCCCACAACAAGCTGGACCTGTACCA     3239
gi|126033139|gb|EF076724.1|       CGTGCCGCTGACCAGCCTGCGAGTGCTCGACCTGTCCCACAACAAGCTGGACCTGTACCA     3239
gi|126033149|gb|EF076729.1|       CGTGCCGCTGACCAGCCTGCGAGTGCTCGACCTGTCCCACAACAAGCTGGACCTGTACCA     3239
gi|126033147|gb|EF076728.1|       CGTGCCGCTGACCAGCCTGCGAGTGCTCGACCTGTCCCACAACAAGCTGGACCTGTACCA     3239
gi|126033137|gb|EF076723.1|       CGTGCCGCTGACCAGCCTGCGAGTGCTCGACCTGTCCCACAACAAGCTGGACCTGTACCA     3239
gi|57471257|gb|AY859726.1|        CGTGCCGCTGACCAGCCTGCGAGTGCTCGACCTGTCCCACAACAAGCTGGACCTGTACCA     1687
gi|27848078|emb|AJ509824.2|       CGTGCCGCTGACCAGCCTGCGAGTGCTCGACCTGTCCCACAACAAGCTGGACCTGTACCA     2967
gi|34147166|ref|NM_183081.1|      CGTGCCGCTGACCAGCCTGCGAGTGCTCGACCTGTCCCACAACAAGCTGGACCTGTACCA     1727
gi|23821159|emb|AJ509825.1|       CGTGCCGCTGACCAGCCTGCGAGTGCTCGACCTGTCCCACAACAAGCTGGACCTGTACCA     1583
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       TGGGCGCTCATTCACGGAGCTGCCACAGCTGGAGGCACTGGACCTCAGCTACAACAGCCA     3284
gi|126033143|gb|EF076726.1|       TGGGCGCTCATTCACGGAGCTGCCACAGCTGGAGGCACTGGACCTCAGCTACAACAGCCA     3299
gi|126033153|gb|EF076731.1|       TGGGCGCTCATTCACGGAGCTGCCACAGCTGGAGGCACTGGACCTCAGCTACAACAGCCA     3299
gi|126033145|gb|EF076727.1|       TGGGCGCTCATTCACGGAGCTGCCACAGCTGGAGGCACTGGACCTCAGCTACAACAGCCA     3299
gi|126033141|gb|EF076725.1|       TGGGCGCTCATTCACGGAGCTGCCACAGCTGGAGGCACTGGACCTCAGCTACAACAGCCA     3299
gi|126033139|gb|EF076724.1|       TGGGCGCTCATTCACGGAGCTGCCACAGCTGGAGGCACTGGACCTCAGCTACAACAGCCA     3299
gi|126033149|gb|EF076729.1|       TGGGCGCTCATTCACGGAGCTGCCACAGCTGGAGGCACTGGACCTCAGCTACAACAGCCA     3299
gi|126033147|gb|EF076728.1|       TGGGCGCTCATTCACGGAGCTGCCACAGCTGGAGGCACTGGACCTCAGCTACAACAGCCA     3299
gi|126033137|gb|EF076723.1|       TGGGCGCTCATTCACGGAGCTGCCACAGCTGGAGGCACTGGACCTCAGCTACAACAGCCA     3299
gi|57471257|gb|AY859726.1|        TGGGCGCTCATTCACGGAGCTGCCGCAGCTGGAGGCACTGGACCTCAGCTACAACAGCCA     1747
gi|27848078|emb|AJ509824.2|       TGGGCGCTCATTCACGGAGCTGCCACAGCTGGAGGCACTGGACCTCAGCTACAACAGCCA     3027
gi|34147166|ref|NM_183081.1|      TGGGCGCTCATTCACGGAGCTGCCACAGCTGGAGGCACTGGACCTCAGCTACAACAGCCA     1787
gi|23821159|emb|AJ509825.1|       TGGGCGCTCATTCACGGAGCTGCCACAGCTGGAGGCACTGGACCTCAGCTACAACAGCCA     1643
                                  ************************ ***********************************

gi|407316560|gb|FJ495080.2|       GCCCTTCAGCATGCAGGGCGTGGGCCACAACCTCAGCTTCGTGGCCCAGCTGCCCTCCCT     3344
gi|126033143|gb|EF076726.1|       GCCCTTCAGCATGCAGGGCGTGGGCCACAACCTCAGCTTCGTGGCCCAGCTGCCCTCCCT     3359
gi|126033153|gb|EF076731.1|       GCCCTTCAGCATGCAGGGCGTGGGCCACAACCTCAGCTTCGTGGCCCAGCTGCCCTCCCT     3359
gi|126033145|gb|EF076727.1|       GCCCTTCAGCATGCAGGGCGTGGGCCACAACCTCAGCTTCGTGGCCCAGCTGCCCTCCCT     3359
gi|126033141|gb|EF076725.1|       GCCCTTCAGCATGCAGGGCGTGGGCCACAACCTCAGCTTCGTGGCCCAGCTGCCCTCCCT     3359
gi|126033139|gb|EF076724.1|       GCCCTTCAGCATGCAGGGCGTGGGCCACAACCTCAGCTTCGTGGCCCAGCTGCCCTCCCT     3359
gi|126033149|gb|EF076729.1|       GCCCTTCAGCATGCAGGGCGTGGGCCACAACCTCAGCTTCGTGGCCCAGCTGCCCTCCCT     3359
gi|126033147|gb|EF076728.1|       GCCCTTCAGCATGCAGGGCGTGGGCCACAACCTCAGCTTCGTGGCCCAGCTGCCCTCCCT     3359
gi|126033137|gb|EF076723.1|       GCCCTTCAGCATGCAGGGCGTGGGCCACAACCTCAGCTTCGTGGCCCAGCTGCCCTCCCT     3359
gi|57471257|gb|AY859726.1|        GCCCTTCAGCATGCAGGGCGTGGGCCACAACCTCAGCTTCGTGGCCCAGCTGCCCTCCCT     1807
gi|27848078|emb|AJ509824.2|       GCCCTTCAGCATGCAGGGCGTGGGCCACAACCTCAGCTTCGTGGCCCAGCTGCCCTCCCT     3087
gi|34147166|ref|NM_183081.1|      GCCCTTCAGCATGCAGGGCGTGGGCCACAACCTCAGCTTCGTGGCCCAGCTGCCCTCCCT     1847
gi|23821159|emb|AJ509825.1|       GCCCTTCAGCATGCAGGGCGTGGGCCACAACCTCAGCTTCGTGGCCCAGCTGCCCTCCCT     1703
```

```
                                 ************************************************************

gi|407316560|gb|FJ495080.2|      GCGCTACCTCAGCCTTGCGCACAATGGCATCCACAGCCGCGTGTCACAGAAGCTCAGCAG     3404
gi|126033143|gb|EF076726.1|      GCGCTACCTCAGCCTTGCGCACAATGGCATCCACAGCCGCGTGTCACAGAAGCTCAGCAG     3419
gi|126033153|gb|EF076731.1|      GCGCTACCTCAGCCTTGCGCACAATGGCATCCACAGCCGCGTGTCACAGAAGCTCAGCAG     3419
gi|126033145|gb|EF076727.1|      GCGCTACCTCAGCCTTGCGCACAATGGCATCCACAGCCGCGTGTCACAGAAGCTCAGCAG     3419
gi|126033141|gb|EF076725.1|      GCGCTACCTCAGCCTTGCGCACAATGGCATCCACAGCCGCGTGTCACAGAAGCTCAGCAG     3419
gi|126033139|gb|EF076724.1|      GCGCTACCTCAGCCTTGCGCACAATGGCATCCACAGCCGCGTGTCACAGAAGCTCAGCAG     3419
gi|126033149|gb|EF076729.1|      GCGCTACCTCAGCCTTGCGCACAATGGCATCCACAGCCGCGTGTCACAGAAGCTCAGCAG     3419
gi|126033147|gb|EF076728.1|      GCGCTACCTCAGCCTTGCGCACAATGGCATCCACAGCCGCGTGTCACAGAAGCTCAGCAG     3419
gi|126033137|gb|EF076723.1|      GCGCTACCTCAGCCTTGCGCACAATGGCATCCACAGCCGCGTGTCACAGAAGCTCAGCAG     3419
gi|57471257|gb|AY859726.1|       GCGCTACCTCAGCCTTGCGCACAATGGCATCCACAGCCGCGTGTCACAGAAGCTCAGCAG     1867
gi|27848078|emb|AJ509824.2|      GCGCTACCTCAGCCTTGCGCACAATGGCATCCACAGCCGCGTGTCACAGAAGCTCAGCAG     3147
gi|34147166|ref|NM_183081.1|     GCGCTACCTCAGCCTTGCGCACAATGGCATCCACAGCCGCGTGTCACAGAAGCTCAGCAG     1907
gi|23821159|emb|AJ509825.1|      GCGCTACCTCAGCCTTGCGCACAATGGCATCCACAGCCGCGTGTCACAGAAGCTCAGCAG     1763
                                 ************************************************************

gi|407316560|gb|FJ495080.2|      CGCCTCGTTGCGCGCCCTGGACTTCAGCGGCAACTCCCTGAGCCAGATGTGGGCCGAGGG     3464
gi|126033143|gb|EF076726.1|      CGCCTCGTTGCGCGCCCTGGACTTCAGCGGCAACTCCCTGAGCCAGATGTGGGCCGAGGG     3479
gi|126033153|gb|EF076731.1|      CGCCTCGTTGCGCGCCCTGGACTTCAGCGGCAACTCCCTGAGCCAGATGTGGGCGGAGGG     3479
gi|126033145|gb|EF076727.1|      CGCCTCGTTGCGCGCCCTGGACTTCAGCGGCAACTCCCTGAGCCAGATGTGGGCGGAGGG     3479
gi|126033141|gb|EF076725.1|      CGCCTCGTTGCGCGCCCTGGACTTCAGCGGCAACTCCCTGAGCCAGATGTGGGCGGAGGG     3479
gi|126033139|gb|EF076724.1|      CGCCTCGTTGCGCGCCCTGGACTTCAGCGGCAACTCCCTGAGCCAGATGTGGGCGGAGGG     3479
gi|126033149|gb|EF076729.1|      CGCCTCGTTGCGCGCCCTGGACTTCAGCGGCAACTCCCTGAGCCAGATGTGGGCGGAGGG     3479
gi|126033147|gb|EF076728.1|      CGCCTCGTTGCGCGCCCTGGACTTCAGCGGCAACTCCCTGAGCCAGATGTGGGCGGAGGG     3479
gi|126033137|gb|EF076723.1|      CGCCTCGTTGCGCGCCCTGGACTTCAGCGGCAACTCCCTGAGCCAGATGTGGGCGGAGGG     3479
gi|57471257|gb|AY859726.1|       CGCCTCGTTGCGCGCCCTGGACTTCAGCGGCAACTCCCTGAGCCAGATGTGGGCCGAGGG     1927
gi|27848078|emb|AJ509824.2|      CGCCTCGTTGCGCGCCCTGGACTTCAGCGGCAACTCCCTGAGCCAGATGTGGGCGGAGGG     3207
gi|34147166|ref|NM_183081.1|     CGCCTCGTTGCGCGCCCTGGACTTCAGCGGCAACTCCCTGAGCCAGATGTGGGCGGAGGG     1967
gi|23821159|emb|AJ509825.1|      CGCCTCGTTGCGCGCCCTGGACTTCAGCGGCAACTCCCTGAGCCAGATGTGGGCGGAGGG     1823
                                 ************************************************** *****

gi|407316560|gb|FJ495080.2|      AGACCTCTATCTCTGCTTTTTCAAAGGCTTGAGGAACCTGGTCCAGCTGGACCTGTCCGA     3524
gi|126033143|gb|EF076726.1|      AGACCTCTATCTCTGCTTTTTCAAAGGCTTGAGGAACCTGGTCCAGCTGGACCTGTCCGA     3539
gi|126033153|gb|EF076731.1|      AGACCTCTATCTCTGCTTTTTCAAAGGCTTGAGGAACCTGGTCCAGCTGGACCTGTCCGA     3539
gi|126033145|gb|EF076727.1|      AGACCTCTATCTCTGCTTTTTCAAAGGCTTGAGGAACCTGGTCCAGCTGGACCTGTCCGA     3539
gi|126033141|gb|EF076725.1|      AGACCTCTATCTCTGCTTTTTCAAAGGCTTGAGGAACCTGGTCCAGCTGGACCTGTCCGA     3539
gi|126033139|gb|EF076724.1|      AGACCTCTATCTCTGCTTTTTCAAAGGCTTGAGGAACCTGGTCCAGCTGGACCTGTCCGA     3539
gi|126033149|gb|EF076729.1|      AGACCTCTATCTCTGCTTTTTCAAAGGCTTGAGGAACCTGGTCCAGCTGGACCTGTCCGA     3539
```

```
gi|126033147|gb|EF076728.1|       AGACCTCTATCTCTGCTTTTTCAAAGGCTTGAGGAACCTGGTCCAGCTGGACCTGTCCGA     3539
gi|126033137|gb|EF076723.1|       AGACCTCTATCTCTGCTTTTTCAAAGGCTTGAGGAACCTGGTCCAGCTGGACCTGTCCGA     3539
gi|57471257|gb|AY859726.1|        AGACCTCTATCTCTGCTTTTTCAAAGGCTTGAGGAACCTGGTCCAGCTGGACCTGTCCGA     1987
gi|27848078|emb|AJ509824.2|       AGACCTCTATCTCTGCTTTTTCAAAGGCTTGAGGAACCTGGTCCAGCTGGACCTGTCCGA     3267
gi|34147166|ref|NM_183081.1|      AGACCTCTATCTCTGCTTTTTCAAAGGCTTGAGGAACCTGGTCCAGCTGGACCTGTCCGA     2027
gi|23821159|emb|AJ509825.1|       AGACCTCTATCTCTGCTTTTTCAAAGGCTTGAGGAACCTGGTCCAGCTGGACCTGTCCGA     1883
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       GAACCATCTGCACACCCTCCTGCCTCGTCACCTGGACAACCTGCCCAAGAGCCTGCGGCA     3584
gi|126033143|gb|EF076726.1|       GAACCATCTGCACACCCTCCTGCCTCGTCACCTGGACAACCTGCCCAAGAGCCTGCGGCA     3599
gi|126033153|gb|EF076731.1|       GAACCATCTGCACACCCTCCTGCCTCGTCACCTGGACAACCTGCCCAAGAGCCTGCGGCA     3599
gi|126033145|gb|EF076727.1|       GAACCATCTGCACACCCTCCTGCCTCGTCACCTGGACAACCTGCCCAAGAGCCTGCGGCA     3599
gi|126033141|gb|EF076725.1|       GAACCATCTGCACACCCTCCTGCCTCGTCACCTGGACAACCTGCCCAAGAGCCTGCGGCA     3599
gi|126033139|gb|EF076724.1|       GAACCATCTGCACACCCTCCTGCCTCGTCACCTGGACAACCTGCCCAAGAGCCTGCGGCA     3599
gi|126033149|gb|EF076729.1|       GAACCATCTGCACACCCTCCTGCCTCGTCACCTGGACAACCTGCCCAAGAGCCTGCGGCA     3599
gi|126033147|gb|EF076728.1|       GAACCATCTGCACACCCTCCTGCCTCGTCACCTGGACAACCTGCCCAAGAGCCTGCGGCA     3599
gi|126033137|gb|EF076723.1|       GAACCATCTGCACACCCTCCTGCCTCGTCACCTGGACAACCTGCCCAAGAGCCTGCGGCA     3599
gi|57471257|gb|AY859726.1|        GAACCATCTGCACACCCTCCTGCCTCGTCACCTGGACAACCTGCCCAAGAGCCTGCGGCA     2047
gi|27848078|emb|AJ509824.2|       GAACCATCTGCACACCCTCCTGCCTCGTCACCTGGACAACCTGCCCAAGAGCCTGCGGCA     3327
gi|34147166|ref|NM_183081.1|      GAACCATCTGCACACCCTCCTGCCTCGTCACCTGGACAACCTGCCCAAGAGCCTGCGGCA     2087
gi|23821159|emb|AJ509825.1|       GAACCATCTGCACACCCTCCTGCCTCGTCACCTGGACAACCTGCCCAAGAGCCTGCGGCA     1943
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       GCTGCGTCTCCGGGACAATAACCTGGCCTTCTTCAACTGGAGCAGCCTGACCGTCCTGCC     3644
gi|126033143|gb|EF076726.1|       GCTGCGTCTCCGGGACAATAACCTGGCCTTCTTCAACTGGAGCAGCCTGACCGTCCTGCC     3659
gi|126033153|gb|EF076731.1|       GCTGCGTCTCCGGGACAATAACCTGGCCTTCTTCAACTGGAGCAGCCTGACCGTCCTGCC     3659
gi|126033145|gb|EF076727.1|       GCTGCGTCTCCGGGACAATAACCTGGCCTTCTTCAACTGGAGCAGCCTGACCGTCCTGCC     3659
gi|126033141|gb|EF076725.1|       GCTGCGTCTCCGGGACAATAACCTGGCCTTCTTCAACTGGAGCAGCCTGACCGTCCTGCC     3659
gi|126033139|gb|EF076724.1|       GCTGCGTCTCCGGGACAATAACCTGGCCTTCTTCAACTGGAGCAGCCTGACCGTCCTGCC     3659
gi|126033149|gb|EF076729.1|       GCTGCGTCTCCGGGACAATAACCTGGCCTTCTTCAACTGGAGCAGCCTGACCGTCCTGCC     3659
gi|126033147|gb|EF076728.1|       GCTGCGTCTCCGGGACAATAACCTGGCCTTCTTCAACTGGAGCAGCCTGACCGTCCTGCC     3659
gi|126033137|gb|EF076723.1|       GCTGCGTCTCCGGGACAATAACCTGGCCTTCTTCAACTGGAGCAGCCTGACCGTCCTGCC     3659
gi|57471257|gb|AY859726.1|        GCTGCGTCTCCGGGACAATAACCTGGCCTTCTTCAACTGGAGCAGCCTGACCGTCCTGCC     2107
gi|27848078|emb|AJ509824.2|       GCTGCGTCTCCGGGACAATAACCTGGCCTTCTTCAACTGGAGCAGCCTGACCGTCCTGCC     3387
gi|34147166|ref|NM_183081.1|      GCTGCGTCTCCGGGACAATAACCTGGCCTTCTTCAACTGGAGCAGCCTGACCGTCCTGCC     2147
gi|23821159|emb|AJ509825.1|       GCTGCGTCTCCGGGACAATAACCTGGCCTTCTTCAACTGGAGCAGCCTGACCGTCCTGCC     2003
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       CCGGCTGGAAGCCCTGGATCTGGCAGGAAACCAGCTGAAGGCCCTGAGCAACGGCAGCCT     3704
```

```
gi|126033143|gb|EF076726.1|       CCGGCTGGAAGCCCTGGATCTGGCAGGAAACCAGCTGAAGGCCCTGAGCAACGGCAGCCT     3719
gi|126033153|gb|EF076731.1|       CCGGCTGGAAGCCCTGGATCTGGCAGGAAACCAGCTGAAGGCCCTGAGCAACGGCAGCCT     3719
gi|126033145|gb|EF076727.1|       CCGGCTGGAAGCCCTGGATCTGGCAGGAAACCAGCTGAAGGCCCTGAGCAACGGCAGCCT     3719
gi|126033141|gb|EF076725.1|       CCGGCTGGAAGCCCTGGATCTGGCAGGAAACCAGCTGAAGGCCCTGAGCAACGGCAGCCT     3719
gi|126033139|gb|EF076724.1|       CCGGCTGGAAGCCCTGGATCTGGCAGGAAACCAGCTGAAGGCCCTGAGCAACGGCAGCCT     3719
gi|126033149|gb|EF076729.1|       CCGGCTGGAAGCCCTGGATCTGGCAGGAAACCAGCTGAAGGCCCTGAGCAACGGCAGCCT     3719
gi|126033147|gb|EF076728.1|       CCGGCTGGAAGCCCTGGATCTGGCAGGAAACCAGCTGAAGGCCCTGAGCAACGGCAGCCT     3719
gi|126033137|gb|EF076723.1|       CCGGCTGGAAGCCCTGGATCTGGCAGGAAACCAGCTGAAGGCCCTGAGCAACGGCAGCCT     3719
gi|57471257|gb|AY859726.1|        CCGGCTGGAAGCCCTGGATCTGGCAGGAAACCAGCTGAAGGCCCTGAGCAACGGCAGCCT     2167
gi|27848078|emb|AJ509824.2|       CCGGCTGGAAGCCCTGGATCTGGCAGGAAACCAGCTGAAGGCCCTGAGCAACGGCAGCCT     3447
gi|34147166|ref|NM_183081.1|      CCGGCTGGAAGCCCTGGATCTGGCAGGAAACCAGCTGAAGGCCCTGAGCAACGGCAGCCT     2207
gi|23821159|emb|AJ509825.1|       CCGGCTGGAAGCCCTGGATCTGGCAGGAAACCAGCTGAAGGCCCTGAGCAACGGCAGCCT     2063
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       GCCGCCTGGCATCCGGCTCCAGAAGCTGGACGTGAGCAGCAACAGCATCGGCTTCGTGAT     3764
gi|126033143|gb|EF076726.1|       GCCGCCTGGCATCCGGCTCCAGAAGCTGGACGTGAGCAGCAACAGCATCGGCTTCGTGAT     3779
gi|126033153|gb|EF076731.1|       GCCGCCTGGCATCCGGCTCCAGAAGCTGGACGTGAGCAGCAACAGCATCGGCTTCGTGAT     3779
gi|126033145|gb|EF076727.1|       GCCGCCTGGCATCCGGCTCCAGAAGCTGGACGTGAGCAGCAACAGCATCGGCTTCGTGAT     3779
gi|126033141|gb|EF076725.1|       GCCGCCTGGCATCCGGCTCCAGAAGCTGGACGTGAGCAGCAACAGCATCGGCTTCGTGAT     3779
gi|126033139|gb|EF076724.1|       GCCGCCTGGCATCCGGCTCCAGAAGCTGGACGTGAGCAGCAACAGCATCGGCTTCGTGAT     3779
gi|126033149|gb|EF076729.1|       GCCGCCTGGCATCCGGCTCCAGAAGCTGGACGTGAGCAGCAACAGCATCGGCTTCGTGAT     3779
gi|126033147|gb|EF076728.1|       GCCGCCTGGCATCCGGCTCCAGAAGCTGGACGTGAGCAGCAACAGCATCGGCTTCGTGAT     3779
gi|126033137|gb|EF076723.1|       GCCGCCTGGCATCCGGCTCCAGAAGCTGGACGTGAGCAGCAACAGCATCGGCTTCGTGAT     3779
gi|57471257|gb|AY859726.1|        GCCGCCTGGCATCCGGCTCCAGAAGCTGGACGTGAGCAGCAACAGCATCGGCTTCGTGAT     2227
gi|27848078|emb|AJ509824.2|       GCCGCCTGGCATCCGGCTCCAGAAGCTGGAGGTGAGCAGCAACAGCATCGGCTTCGTGAT     3507
gi|34147166|ref|NM_183081.1|      GCCGCCTGGCATCCGGCTCCAGAAGCTGGAGGTGAGCAGCAACAGCATCGGCTTCGTGAT     2267
gi|23821159|emb|AJ509825.1|       GCCGCCTGGCATCCGGCTCCAGAAGCTGGAGGTGAGCAGCAACAGCATCGGCTTCGTGAT     2123
                                  ****************************** ******************************

gi|407316560|gb|FJ495080.2|       CCCCGGCTTCTTCGTCCGCGCGACTCGGCTGATAGAGCTTAACCTCAGCGCCAATGCCCT     3824
gi|126033143|gb|EF076726.1|       CCCCGGCTTCTTCGTCCGCGCGACTCGGCTGATAGAGCTTAACCTCAGCGCCAATGCCCT     3839
gi|126033153|gb|EF076731.1|       CCCCGGCTTCTTCGTCCGCGCGACTCGGCTGATAGAGCTTAACCTCAGCGCCAATGCCCT     3839
gi|126033145|gb|EF076727.1|       CCCCGGCTTCTTCGTCCGCGCGACTCGGCTGATAGAGCTTAACCTCAGCGCCAATGCCCT     3839
gi|126033141|gb|EF076725.1|       CCCCGGCTTCTTCGTCCGCGCGACTCGGCTGATAGAGCTTAACCTCAGCGCCAATGCCCT     3839
gi|126033139|gb|EF076724.1|       CCCCGGCTTCTTCGTCCGCGCGACTCGGCTGATAGAGCTTAACCTCAGCGCCAATGCCCT     3839
gi|126033149|gb|EF076729.1|       CCCCGGCTTCTTCGTCCGCGCGACTCGGCTGATAGAGCTTAACCTCAGCGCCAATGCCCT     3839
gi|126033147|gb|EF076728.1|       CCCCGGCTTCTTCGTCCGCGCGACTCGGCTGATAGAGCTTAACCTCAGCGCCAATGCCCT     3839
gi|126033137|gb|EF076723.1|       CCCCGGCTTCTTCGTCCGCGCGACTCGGCTGATAGAGCTTAACCTCAGCGCCAATGCCCT     3839
gi|57471257|gb|AY859726.1|        CCCCGGCTTCTTCGTCCGCGCGACTCGGCTGATAGAGCTTAACCTCAGCGCCAATGCCCT     2287
```

```
gi|27848078|emb|AJ509824.2|       CCCCGGCTTCTTCGTCCGCGCGACTCGGCTGATAGAGCTTAACCTCAGCGCCAATGCCCT     3567
gi|34147166|ref|NM_183081.1|      CCCCGGCTTCTTCGTCCGCGCGACTCGGCTGATAGAGCTTAACCTCAGCGCCAATGCCCT     2327
gi|23821159|emb|AJ509825.1|       CCCCGGCTTCTTCGTCCGCGCGACTCGGCTGATAGAGCTTAACCTCAGCGCCAATGCCCT     2183
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       GAAGACAGTGGATCCCTCCTGGTTCGGTTCCTTAGCAGGGACCCTGAAAATCCTAGACGT     3884
gi|126033143|gb|EF076726.1|       GAAGACAGTGGATCCCTCCTGGTTCGGTTCCTTAGCAGGGACCCTGAAAATCCTAGACGT     3899
gi|126033153|gb|EF076731.1|       GAAGACAGTGGATCCCTCCTGGTTCGGTTCCTTAGCAGGGACCCTGAAAATCCTAGACGT     3899
gi|126033145|gb|EF076727.1|       GAAGACAGTGGATCCCTCCTGGTTCGGTTCCTTAGCAGGGACCCTGAAAATCCTAGACGT     3899
gi|126033141|gb|EF076725.1|       GAAGACAGTGGATCCCTCCTGGTTCGGTTCCTTAGCAGGGACCCTGAAAATCCTAGACGT     3899
gi|126033139|gb|EF076724.1|       GAAGACAGTGGATCCCTCCTGGTTCGGTTCCTTAGCAGGGACCCTGAAAATCCTAGACGT     3899
gi|126033149|gb|EF076729.1|       GAAGACAGTGGATCCCTCCTGGTTCGGTTCCTTAGCAGGGACCCTGAAAATCCTAGACGT     3899
gi|126033147|gb|EF076728.1|       GAAGACAGTGGATCCCTCCTGGTTCGGTTCCTTAGCAGGGACCCTGAAAATCCTAGACGT     3899
gi|126033137|gb|EF076723.1|       GAAGACAGTGGATCCCTCCTGGTTCGGTTCCTTAGCAGGGACCCTGAAAATCCTAGACGT     3899
gi|57471257|gb|AY859726.1|        GAAGACAGTGGATCCCTCCTGGTTCGGTTCCTTAGCAGGGACCCTGAAAATCCTAGACGT     2347
gi|27848078|emb|AJ509824.2|       GAAGACAGTGGATCCCTCCTGGTTCGGTTCCTTAGCAGGGACCCTGAAAATCCTAGACGT     3627
gi|34147166|ref|NM_183081.1|      GAAGACAGTGGATCCCTCCTGGTTCGGTTCCTTAGCAGGGACCCTGAAAATCCTAGACGT     2387
gi|23821159|emb|AJ509825.1|       GAAGACAGTGGATCCCTCCTGGTTCGGTTCCTTAGCAGGGACCCTGAAAATCCTAGACGT     2243
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       GAGCGCCAACCCGCTCCACTGCGCCTGCGGGGCGGCCTTTGTGGACTTCCTGCTGGAGAG     3944
gi|126033143|gb|EF076726.1|       GAGCGCCAACCCGCTCCACTGCGCCTGCGGGGCGGCCTTTGTGGACTTCCTGCTGGAGAG     3959
gi|126033153|gb|EF076731.1|       GAGCGCCAACCCGCTCCACTGCGCCTGCGGGGCGGCCTTTGTGGACTTCCTGCTGGAGAG     3959
gi|126033145|gb|EF076727.1|       GAGCGCCAACCCGCTCCACTGCGCCTGCGGGGCGGCCTTTGTGGACTTCCTGCTGGAGAG     3959
gi|126033141|gb|EF076725.1|       GAGCGCCAACCCGCTCCACTGCGCCTGCGGGGCGGCCTTTGTGGACTTCCTGCTGGAGAG     3959
gi|126033139|gb|EF076724.1|       GAGCGCCAACCCGCTCCACTGCGCCTGCGGGGCGGCCTTTGTGGACTTCCTGCTGGAGAG     3959
gi|126033149|gb|EF076729.1|       GAGCGCCAACCCGCTCCACTGCGCCTGCGGGGCGGCCTTTGTGGACTTCCTGCTGGAGAG     3959
gi|126033147|gb|EF076728.1|       GAGCGCCAACCCGCTCCACTGCGCCTGCGGGGCGGCCTTTGTGGACTTCCTGCTGGAGAG     3959
gi|126033137|gb|EF076723.1|       GAGCGCCAACCCGCTCCACTGCGCCTGCGGGGCGGCCTTTGTGGACTTCCTGCTGGAGAG     3959
gi|57471257|gb|AY859726.1|        GAGCGCCAACCCGCTCCACTGCGCCTGCGGGGCGGCCTTTGTGGACTTCCTGCTGGAGAG     2407
gi|27848078|emb|AJ509824.2|       GAGCGCCAACCCGCTCCACTGCGCCTGCGGGGCGGCCTTTGTGGACTTCCTGCTGGAGAG     3687
gi|34147166|ref|NM_183081.1|      GAGCGCCAACCCGCTCCACTGCGCCTGCGGGGCGGCCTTTGTGGACTTCCTGCTGGAGAG     2447
gi|23821159|emb|AJ509825.1|       GAGCGCCAACCCGCTCCACTGCGCCTGCGGGGCGGCCTTTGTGGACTTCCTGCTGGAGAG     2303
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       ACAGGAGGCCGTGCCCGGGCTGTCCAGGCGCGTCACATGTGGCAGTCCGGGCCAGCTCCA     4004
gi|126033143|gb|EF076726.1|       ACAGGAGGCCGTGCCCGGGCTGTCCAGGCGCGTCACATGTGGCAGTCCGGGCCAGCTCCA     4019
gi|126033153|gb|EF076731.1|       ACAGGAGGCCGTGCCCGGGCTGTCCAGGCGCGTCACATGTGGCAGTCCGGGCCAGCTCCA     4019
gi|126033145|gb|EF076727.1|       ACAGGAGGCCGTGCCCGGGCTGTCCAGGCGCGTCACATGTGGCAGTCCGGGCCAGCTCCA     4019
```

```
gi|126033141|gb|EF076725.1|       ACAGGAGGCCGTGCCCGGGCTGTCCAGGCGCGTCACATGTGGCAGTCCGGGCCAGCTCCA     4019
gi|126033139|gb|EF076724.1|       ACAGGAGGCCGTGCCCGGGCTGTCCAGGCGCGTCACATGTGGCAGTCCGGGCCAGCTCCA     4019
gi|126033149|gb|EF076729.1|       ACAGGAGGCCGTGCCCGGGCTGTCCAGGCGCGTCACATGTGGCAGTCCGGGCCAGCTCCA     4019
gi|126033147|gb|EF076728.1|       ACAGGAGGCCGTGCCCGGGCTGTCCAGGCGCGTCACATGTGGCAGTCCGGGCCAGCTCCA     4019
gi|126033137|gb|EF076723.1|       ACAGGAGGCCGTGCCCGGGCTGTCCAGGCGCGTCACATGTGGCAGTCCGGGCCAGCTCCA     4019
gi|57471257|gb|AY859726.1|        ACAGGAGGCCGTGCCCGGGCTGTCCAGGCGCGTCACATGTGGCAGTCCGGGCCAGCTCCA     2467
gi|27848078|emb|AJ509824.2|       ACAGGAGGCCGTGCCCGGGCTGTCCAGGCGCGTCACATGTGGCAGTCCGGGCCAGCTCCA     3747
gi|34147166|ref|NM_183081.1|      ACAGGAGGCCGTGCCCGGGCTGTCCAGGCGCGTCACATGTGGCAGTCCGGGCCAGCTCCA     2507
gi|23821159|emb|AJ509825.1|       ACAGGAGGCCGTGCCCGGGCTGTCCAGGCGCGTCACATGTGGCAGTCCGGGCCAGCTCCA     2363
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       GGGCCGCAGCATCTTCACACAGGACCTGCGTCTCTGCCTGGATGAGACCCTCTCCTTGGA     4064
gi|126033143|gb|EF076726.1|       GGGCCGCAGCATCTTCACACAGGACCTGCGTCTCTGCCTGGATGAGACCCTCTCCTTGGA     4079
gi|126033153|gb|EF076731.1|       GGGCCGCAGCATCTTCACACAGGACCTGCGTCTCTGCCTGGATGAGACCCTCTCCTTGGA     4079
gi|126033145|gb|EF076727.1|       GGGCCGCAGCATCTTCACACAGGACCTGCGTCTCTGCCTGGATGAGACCCTCTCCTTGGA     4079
gi|126033141|gb|EF076725.1|       GGGCCGCAGCATCTTCACACAGGACCTGCGTCTCTGCCTGGATGAGACCCTCTCCTTGGA     4079
gi|126033139|gb|EF076724.1|       GGGCCGCAGCATCTTCACACAGGACCTGCGTCTCTGCCTGGATGAGACCCTCTCCTTGGA     4079
gi|126033149|gb|EF076729.1|       GGGCCGCAGCATCTTCACACAGGACCTGCGTCTCTGCCTGGATGAGACCCTCTCCTTGGA     4079
gi|126033147|gb|EF076728.1|       GGGCCGCAGCATCTTCACACAGGACCTGCGTCTCTGCCTGGATGAGACCCTCTCCTTGGA     4079
gi|126033137|gb|EF076723.1|       GGGCCGCAGCATCTTCACACAGGACCTGCGTCTCTGCCTGGATGAGACCCTCTCCTTGGA     4079
gi|57471257|gb|AY859726.1|        GGGCCGCAGCATCTTCACACAGGACCTGCGCCTCTGCCTGGATGAGACCCTCTCCTTGGA     2527
gi|27848078|emb|AJ509824.2|       GGGCCGCAGCATCTTCACACAGGACCTGCGCCTCTGCCTGGATGAGACCCTCTCCTTGGA     3807
gi|34147166|ref|NM_183081.1|      GGGCCGCAGCATCTTCACACAGGACCTGCGCCTCTGCCTGGATGAGACCCTCTCCTTGGA     2567
gi|23821159|emb|AJ509825.1|       GGGCCGCAGCATCTTCACACAGGACCTGCGCCTCTGCCTGGATGAGACCCTCTCCTTGGA     2423
                                  ****************************** ******************************

gi|407316560|gb|FJ495080.2|       CTGCTTTGGCCTCTCGCTGCTAATGGTGGCGCTGGGCCTGGCAGTGCCCATGCTGCACCA     4124
gi|126033143|gb|EF076726.1|       CTGCTTTGGCCTCTCGCTGCTAATGGTGGCGCTGGGCCTGGCAGTGCCCATGCTGCACCA     4139
gi|126033153|gb|EF076731.1|       CTGCTTTGGCCTCTCGCTGCTAATGGTGGCGCTGGGCCTGGCAGTGCCCATGCTGCACCA     4139
gi|126033145|gb|EF076727.1|       CTGCTTTGGCCTCTCGCTGCTAATGGTGGCGCTGGGCCTGGCAGTGCCCATGCTGCACCA     4139
gi|126033141|gb|EF076725.1|       CTGCTTTGGCCTCTCGCTGCTAATGGTGGCGCTGGGCCTGGCAGTGCCCATGCTGCACCA     4139
gi|126033139|gb|EF076724.1|       CTGCTTTGGCCTCTCGCTGCTAATGGTGGCGCTGGGCCTGGCAGTGCCCATGCTGCACCA     4139
gi|126033149|gb|EF076729.1|       CTGCTTTGGCCTCTCGCTGCTAATGGTGGCGCTGGGCCTGGCAGTGCCCATGCTGCACCA     4139
gi|126033147|gb|EF076728.1|       CTGCTTTGGCCTCTCGCTGCTAATGGTGGCGCTGGGCCTGGCAGTGCCCATGCTGCACCA     4139
gi|126033137|gb|EF076723.1|       CTGCTTTGGCCTCTCGCTGCTAATGGTGGCGCTGGGCCTGGCAGTGCCCATGCTGCACCA     4139
gi|57471257|gb|AY859726.1|        CTGCTTTGGCCTCTCACTGCTAATGGTGGCGCTGGGCCTGGCAGTGCCCATGCTGCACCA     2587
gi|27848078|emb|AJ509824.2|       CTGCTTTGGCCTCTCGCTGCTAATGGTGGCGCTGGGCCTGGCAGTGCCCATGCTGCACCA     3867
gi|34147166|ref|NM_183081.1|      CTGCTTTGGCCTCTCGCTGCTAATGGTGGCGCTGGGCCTGGCAGTGCCCATGCTGCACCA     2627
gi|23821159|emb|AJ509825.1|       CTGCTTTGGCCTCTCGCTGCTAATGGTGGCGCTGGGCCTGGCAGTGCCCATGCTGCACCA     2483
```

```
                          *************** ******************************************

gi|407316560|gb|FJ495080.2|       CCTCTGTGGCTGGGACCTCTGGTACTGCTTCCACCTGTGTCTGGCCCATTTGCCCCGACG     4184
gi|126033143|gb|EF076726.1|       CCTCTGTGGCTGGGACCTCTGGTACTGCTTCCACCTGTGTCTGGCCCATTTGCCCCGACG     4199
gi|126033153|gb|EF076731.1|       CCTCTGTGGCTGGGACCTCTGGTACTGCTTCCACCTGTGTCTGGCCCATTTGCCCCGACG     4199
gi|126033145|gb|EF076727.1|       CCTCTGTGGCTGGGACCTCTGGTACTGCTTCCACCTGTGTCTGGCCCATTTGCCCCGACG     4199
gi|126033141|gb|EF076725.1|       CCTCTGTGGCTGGGACCTCTGGTACTGCTTCCACCTGTGTCTGGCCCATTTGCCCCGACG     4199
gi|126033139|gb|EF076724.1|       CCTCTGTGGCTGGGACCTCTGGTACTGCTTCCACCTGTGTCTGGCCCATTTGCCCCGACG     4199
gi|126033149|gb|EF076729.1|       CCTCTGTGGCTGGGACCTCTGGTACTGCTTCCACCTGTGTCTGGCCCATTTGCCCCGACG     4199
gi|126033147|gb|EF076728.1|       CCTCTGTGGCTGGGACCTCTGGTACTGCTTCCACCTGTGTCTGGCCCATTTGCCCCGACG     4199
gi|126033137|gb|EF076723.1|       CCTCTGTGGCTGGGACCTCTGGTACTGCTTCCACCTGTGTCTGGCCCATTTGCCCCGACG     4199
gi|57471257|gb|AY859726.1|        CCTCTGTGGCTGGGACCTCTGGTACTGCTTCCACCTGTGTCTGGCCCATTTGCCCCGACG     2647
gi|27848078|emb|AJ509824.2|       CCTCTGTGGCTGGGACCTCTGGTACTGCTTCCACCTGTGTCTGGCCCATTTGCCCCGACG     3927
gi|34147166|ref|NM_183081.1|      CCTCTGTGGCTGGGACCTCTGGTACTGCTTCCACCTGTGTCTGGCCCATTTGCCCCGACG     2687
gi|23821159|emb|AJ509825.1|       CCTCTGTGGCTGGGACCTCTGGTACTGCTTCCACCTGTGTCTGGCCCATTTGCCCCGACG     2543
                          ************************************************************

gi|407316560|gb|FJ495080.2|       GCGGCGGCAGCGGGGCGAGGACACCCTGCTCTATGATGCCTTCGTGGTCTTCGACAAGGT     4244
gi|126033143|gb|EF076726.1|       GCGGCGGCAGCGGGGCGAGGACACCCTGCTCTATGATGCCTTCGTGGTCTTCGACAAGGT     4259
gi|126033153|gb|EF076731.1|       GCGGCGGCAGCGGGGCGAGGACACCCTGCTCTATGATGCCTTCGTGGTCTTCGACAAGGT     4259
gi|126033145|gb|EF076727.1|       GCGGCGGCAGCGGGGCGAGGACACCCTGCTCTATGATGCCTTCGTGGTCTTCGACAAGGT     4259
gi|126033141|gb|EF076725.1|       GCGGCGGCAGCGGGGCGAGGACACCCTGCTCTATGATGCCTTCGTGGTCTTCGACAAGGT     4259
gi|126033139|gb|EF076724.1|       GCGGCGGCAGCGGGGCGAGGACACCCTGCTCTATGATGCCTTCGTGGTCTTCGACAAGGT     4259
gi|126033149|gb|EF076729.1|       GCGGCGGCAGCGGGGCGAGGACACCCTGCTCTATGATGCCTTCGTGGTCTTCGACAAGGT     4259
gi|126033147|gb|EF076728.1|       GCGGCGGCAGCGGGGCGAGGACACCCTGCTCTATGATGCCTTCGTGGTCTTCGACAAGGT     4259
gi|126033137|gb|EF076723.1|       GCGGCGGCAGCGGGGCGAGGACACCCTGCTCTATGATGCCTTCGTGGTCTTCGACAAGGT     4259
gi|57471257|gb|AY859726.1|        GCGGCGGCAGCGGGGCGAGGACACCCTGCTCTATGATGCCGTCGTGGTCTTCGACAAGGT     2707
gi|27848078|emb|AJ509824.2|       GCGGCGGCAGCGGGGCGAGGACACCCTGCTCTATGATGCCTTCGTGGTCTTCGACAAGGT     3987
gi|34147166|ref|NM_183081.1|      GCGGCGGCAGCGGGGCGAGGACACCCTGCTCTATGATGCCTTCGTGGTCTTCGACAAGGT     2747
gi|23821159|emb|AJ509825.1|       GCGGCGGCAGCGGGGCGAGGACACCCTGCTCTATGATGCCTTCGTGGTCTTCGACAAGGT     2603
                          ************************************** ******************

gi|407316560|gb|FJ495080.2|       GCAGAGTGCAGTGGCTGATTGGGTGTACAACGAGCTCCGCGTGCAGCTGGAGGAGCGCCG     4304
gi|126033143|gb|EF076726.1|       GCAGAGTGCAGTGGCTGATTGGGTGTACAACGAGCTCCGCGTGCAGCTGGAGGAGCGCCG     4319
gi|126033153|gb|EF076731.1|       GCAGAGTGCAGTGGCTGATTGGGTGTACAACGAGCTCCGCGTGCAGCTGGAGGAGCGCCG     4319
gi|126033145|gb|EF076727.1|       GCAGAGTGCAGTGGCTGATTGGGTGTACAACGAGCTCCGCGTGCAGCTGGAGGAGCGCCG     4319
gi|126033141|gb|EF076725.1|       GCAGAGTGCAGTGGCTGATTGGGTGTACAACGAGCTCCGCGTGCAGCTGGAGGAGCGCCG     4319
gi|126033139|gb|EF076724.1|       GCAGAGTGCAGTGGCTGATTGGGTGTACAACGAGCTCCGCGTGCAGCTGGAGGAGCGCCG     4319
gi|126033149|gb|EF076729.1|       GCAGAGTGCAGTGGCTGATTGGGTGTACAACGAGCTCCGCGTGCAGCTGGAGGAGCGCCG     4319
```

```
gi|126033147|gb|EF076728.1|       GCAGAGTGCAGTGGCTGATTGGGTGTACAACGAGCTCCGCGTGCAGCTGGAGGAGCGCCG     4319
gi|126033137|gb|EF076723.1|       GCAGAGTGCAGTGGCTGATTGGGTGTACAACGAGCTCCGCGTGCAGCTGGAGGAGCGCCG     4319
gi|57471257|gb|AY859726.1|        GCAGAGTGCAGTGGCTGATTGGGTGTACAACGAGCTCCGCGTGCAGCTGGAGGAGCGCCG     2767
gi|27848078|emb|AJ509824.2|       GCAGAGTGCAGTGGCTGATTGGGTGTACAACGAGCTCCGCGTGCAGCTGGAGGAGCGCCG     4047
gi|34147166|ref|NM_183081.1|      GCAGAGTGCAGTGGCTGATTGGGTGTACAACGAGCTCCGCGTGCAGCTGGAGGAGCGCCG     2807
gi|23821159|emb|AJ509825.1|       GCAGAGTGCAGTGGCTGATTGGGTGTACAACGAGCTCCGCGTGCAGCTGGAGGAGCGCCG     2663
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       GGGGCGCCGGGCGCTCCGCCTCTGCCTGGAGGAGCGAGACTGGCTCCCTGGTAAGACGCT     4364
gi|126033143|gb|EF076726.1|       GGGGCGCCGGGCGCTCCGCCTCTGCCTGGAGGAGCGAGACTGGCTCCCTGGTAAGACGCT     4379
gi|126033153|gb|EF076731.1|       GGGGCGCCGGGCGCTCCGCCTCTGCCTGGAGGAGCGAGACTGGCTCCCTGGTAAGACGCT     4379
gi|126033145|gb|EF076727.1|       GGGGCGCCGGGCGCTCCGCCTCTGCCTGGAGGAGCGAGACTGGCTCCCTGGTAAGACGCT     4379
gi|126033141|gb|EF076725.1|       GGGGCGCCGGGCGCTCCGCCTCTGCCTGGAGGAGCGAGACTGGCTCCCTGGTAAGACGCT     4379
gi|126033139|gb|EF076724.1|       GGGGCGCCGGGCGCTCCGCCTCTGCCTGGAGGAGCGAGACTGGCTCCCTGGTAAGACGCT     4379
gi|126033149|gb|EF076729.1|       GGGGCGCCGGGCGCTCCGCCTCTGCCTGGAGGAGCGAGACTGGCTCCCTGGTAAGACGCT     4379
gi|126033147|gb|EF076728.1|       GGGGCGCCGGGCGCTCCGCCTCTGCCTGGAGGAGCGAGACTGGCTCCCTGGTAAGACGCT     4379
gi|126033137|gb|EF076723.1|       GGGGCGCCGGGCGCTCCGCCTCTGCCTGGAGGAGCGAGACTGGCTCCCTGGTAAGACGCT     4379
gi|57471257|gb|AY859726.1|        GGGGCGCCGGGCGCTCCGCCTCTGCCTGGAGGAGCGAGACTGGCTCCCTGGTAAGACGCT     2827
gi|27848078|emb|AJ509824.2|       GGGGCGCCGGGCGCTCCGCCTCTGCCTGGAGGAGCGAGACTGGCTCCCTGGTAAGACGCT     4107
gi|34147166|ref|NM_183081.1|      GGGGCGCCGGGCGCTCCGCCTCTGCCTGGAGGAGCGAGACTGGCTCCCTGGTAAGACGCT     2867
gi|23821159|emb|AJ509825.1|       GGGGCGCCGGGCGCTCCGCCTCTGCCTGGAGGAGCGAGACTGGCTCCCTGGTAAGACGCT     2723
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       CTTCGAGAACCTGTGGGCCTCGGTCTACAGCAGCCGCAAGACCATGTTCGTGCTGGACCA     4424
gi|126033143|gb|EF076726.1|       CTTCGAGAACCTGTGGGCCTCGGTCTACAGCAGCCGCAAGACCATGTTCGTGCTGGACCA     4439
gi|126033153|gb|EF076731.1|       CTTCGAGAACCTGTGGGCCTCGGTCTACAGCAGCCGCAAGACCATGTTCGTGCTGGACCA     4439
gi|126033145|gb|EF076727.1|       CTTCGAGAACCTGTGGGCCTCGGTCTACAGCAGCCGCAAGACCATGTTCGTGCTGGACCA     4439
gi|126033141|gb|EF076725.1|       CTTCGAGAACCTGTGGGCCTCGGTCTACAGCAGCCGCAAGACCATGTTCGTGCTGGACCA     4439
gi|126033139|gb|EF076724.1|       CTTCGAGAACCTGTGGGCCTCGGTCTACAGCAGCCGCAAGACCATGTTCGTGCTGGACCA     4439
gi|126033149|gb|EF076729.1|       CTTCGAGAACCTGTGGGCCTCGGTCTACAGCAGCCGCAAGACCATGTTCGTGCTGGACCA     4439
gi|126033147|gb|EF076728.1|       CTTCGAGAACCTGTGGGCCTCGGTCTACAGCAGCCGCAAGACCATGTTCGTGCTGGACCA     4439
gi|126033137|gb|EF076723.1|       CTTCGAGAACCTGTGGGCCTCGGTCTACAGCAGCCGCAAGACCATGTTCGTGCTGGACCA     4439
gi|57471257|gb|AY859726.1|        CTTCGAGAACCTGTGGGCCTCGGTCTACAGCAGCCGCAAGACCATGTTCGTGCTGGACCA     2887
gi|27848078|emb|AJ509824.2|       CTTCGAGAACCTGTGGGCCTCGGTCTACAGCAGCCGCAAGACCATGTTTGTGCTGGCCCA     4167
gi|34147166|ref|NM_183081.1|      CTTCGAGAACCTGTGGGCCTCGGTCTACAGCAGCCGCAAGACCATGTTTGTGCTGGCCCA     2927
gi|23821159|emb|AJ509825.1|       CTTCGAGAACCTGTGGGCCTCGGTCTACAGCAGCCGCAAGACCATGTTTGTGCTGGCCCA     2783
                                  ********************************************* ******* ***

gi|407316560|gb|FJ495080.2|       CACGGACCGGGTCAGCGGCCTCCTGCGCGCCAGCTTCCTGCTGGCCCAGCAGCGCCTGTT     4484
```

```
gi|126033143|gb|EF076726.1|       CACGGACCGGGTCAGCGGCCTCCTGCGCGCCAGCTTCCTGCTGGCCCAGCAGCGCCTGTT     4499
gi|126033153|gb|EF076731.1|       CACGGACCGGGTCAGCGGCCTCCTGCGCGCCAGCTTCCTGCTGGCCCAGCAGCGCCTGTT     4499
gi|126033145|gb|EF076727.1|       CACGGACCGGGTCAGCGGCCTCCTGCGCGCCAGCTTCCTGCTGGCCCAGCAGCGCCTGTT     4499
gi|126033141|gb|EF076725.1|       CACGGACCGGGTCAGCGGCCTCCTGCGCGCCAGCTTCCTGCTGGCCCAGCAGCGCCTGTT     4499
gi|126033139|gb|EF076724.1|       CACGGACCGGGTCAGCGGCCTCCTGCGCGCCAGCTTCCTGCTGGCCCAGCAGCGCCTGTT     4499
gi|126033149|gb|EF076729.1|       CACGGACCGGGTCAGCGGCCTCCTGCGCGCCAGCTTCCTGCTGGCCCAGCAGCGCCTGTT     4499
gi|126033147|gb|EF076728.1|       CACGGACCGGGTCAGCGGCCTCCTGCGCGCCAGCTTCCTGCTGGCCCAGCAGCGCCTGTT     4499
gi|126033137|gb|EF076723.1|       CACGGACCGGGTCAGCGGCCTCCTGCGCGCCAGCTTCCTGCTGGCCCAGCAGCGCCTGTT     4499
gi|57471257|gb|AY859726.1|        CACGGACCGGGTCAGCGGCCTCCTGCGCGCCAGCTTCCTGCTGGCCCAGCAGCGCCTGTT     2947
gi|27848078|emb|AJ509824.2|       CACGGACCGGGTCAGCGGCCTCCTGCGCGCCAGCTTCCTGCTGGCCCAGCAGCGCCTGTT     4227
gi|34147166|ref|NM_183081.1|      CACGGACCGGGTCAGCGGCCTCCTGCGCGCCAGCTTCCTGCTGGCCCAGCAGCGCCTGTT     2987
gi|23821159|emb|AJ509825.1|       CACGGACCGGGTCAGCGGCCTCCTGCGCGCCAGCTTCCTGCTGGCCCAGCAGCGCCTGTT     2843
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       GGAGGACCGCAAGGACGTCGTAGTGCTGGTGATCCTGCGCCCCGCCGCCTATCGGTCCCG     4544
gi|126033143|gb|EF076726.1|       GGAGGACCGCAAGGACGTCGTAGTGCTGGTGATCCTGCGCCCCGCCGCCTATCGGTCCCG     4559
gi|126033153|gb|EF076731.1|       GGAGGACCGCAAGGACGTCGTAGTGCTGGTGATCCTGCGCCCCGCCGCCTATCGGTCCCG     4559
gi|126033145|gb|EF076727.1|       GGAGGACCGCAAGGACGTCGTAGTGCTGGTGATCCTGCGCCCCGCCGCCTATCGGTCCCG     4559
gi|126033141|gb|EF076725.1|       GGAGGACCGCAAGGACGTCGTAGTGCTGGTGATCCTGCGCCCCGCCGCCTATCGGTCCCG     4559
gi|126033139|gb|EF076724.1|       GGAGGACCGCAAGGACGTCGTAGTGCTGGTGATCCTGCGCCCCGCCGCCTATCGGTCCCG     4559
gi|126033149|gb|EF076729.1|       GGAGGACCGCAAGGACGTCGTAGTGCTGGTGATCCTGCGCCCCGCCGCCTATCGGTCCCG     4559
gi|126033147|gb|EF076728.1|       GGAGGACCGCAAGGACGTCGTAGTGCTGGTGATCCTGCGCCCCGCCGCCTATCGGTCCCG     4559
gi|126033137|gb|EF076723.1|       GGAGGACCGCAAGGACGTCGTAGTGCTGGTGATCCTGCGCCCCGCCGCCTATCGGTCCCG     4559
gi|57471257|gb|AY859726.1|        GGAGGACCGCAAGGACGTCGTAGTGCTGGTGATCCTGCGCCCCGCCGCCTATCGGTCCCG     3007
gi|27848078|emb|AJ509824.2|       GGAGGACCGCAAGGACGTCGTAGTGCTGGTGATCCTGCGCCCCGCCGCCTATCGGTCCCG     4287
gi|34147166|ref|NM_183081.1|      GGAGGACCGCAAGGACGTCGTAGTGCTGGTGATCCTGCGCCCCGCCGCCTATCGGTCCCG     3047
gi|23821159|emb|AJ509825.1|       GGAGGACCGCAAGGACGTCGTAGTGCTGGTGATCCTGCGCCCCGCCGCCTATCGGTCCCG     2903
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       CTACGTGCGGCTGCGCCAGCGCCTCTGCCGCCAGAGCGTCCTCCTCTGGCCCCACCAGCC     4604
gi|126033143|gb|EF076726.1|       CTACGTGCGGCTGCGCCAGCGCCTCTGCCGCCAGAGCGTCCTCCTCTGGCCCCACCAGCC     4619
gi|126033153|gb|EF076731.1|       CTACGTGCGGCTGCGCCAGCGCCTCTGCCGCCAGAGCGTCCTCCTCTGGCCCCACCAGCC     4619
gi|126033145|gb|EF076727.1|       CTACGTGCGGCTGCGCCAGCGCCTCTGCCGCCAGAGCGTCCTCCTCTGGCCCCACCAGCC     4619
gi|126033141|gb|EF076725.1|       CTACGTGCGGCTGCGCCAGCGCCTCTGCCGCCAGAGCGTCCTCCTCTGGCCCCACCAGCC     4619
gi|126033139|gb|EF076724.1|       CTACGTGCGGCTGCGCCAGCGCCTCTGCCGCCAGAGCGTCCTCCTCTGGCCCCACCAGCC     4619
gi|126033149|gb|EF076729.1|       CTACGTGCGGCTGCGCCAGCGCCTCTGCCGCCAGAGCGTCCTCCTCTGGCCCCACCAGCC     4619
gi|126033147|gb|EF076728.1|       CTACGTGCGGCTGCGCCAGCGCCTCTGCCGCCAGAGCGTCCTCCTCTGGCCCCACCAGCC     4619
gi|126033137|gb|EF076723.1|       CTACGTGCGGCTGCGCCAGCGCCTCTGCCGCCAGAGCGTCCTCCTCTGGCCCCACCAGCC     4619
gi|57471257|gb|AY859726.1|        CTACGTGCGGCTGCGCCAGCGCCTCTGCCGCCAGAGCGTCCTCCTCTGGCCCCACCAGCC     3067
```

```
gi|27848078|emb|AJ509824.2|       CTACGTGCGGCTGCGCCAGCGCCTCTGCCGCCAGAGCGTCCTCCTCTGGCCCCACCAGCC     4347
gi|34147166|ref|NM_183081.1|      CTACGTGCGGCTGCGCCAGCGCCTCTGCCGCCAGAGCGTCCTCCTCTGGCCCCACCAGCC     3107
gi|23821159|emb|AJ509825.1|       CTACGTGCGGCTGCGCCAGCGCCTCTGCCGCCAGAGCGTCCTCCTCTGGCCCCACCAGCC     2963
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       CAGTGGCCAGGGTAGTTTCTGGGCCAACCTGGGCATAGCCCTGACCAGGGACAACCGTCA     4664
gi|126033143|gb|EF076726.1|       CAGTGGCCAGGGTAGTTTCTGGGCCAACCTGGGCATAGCCCTGACCAGGGACAACCGTCA     4679
gi|126033153|gb|EF076731.1|       CAGTGGCCAGGGTAGTTTCTGGGCCAACCTGGGCATAGCCCTGACCAGGGACAACCGTCA     4679
gi|126033145|gb|EF076727.1|       CAGTGGCCAGGGTAGTTTCTGGGCCAACCTGGGCATAGCCCTGACCAGGGACAACCGTCA     4679
gi|126033141|gb|EF076725.1|       CAGTGGCCAGGGTAGTTTCTGGGCCAACCTGGGCATAGCCCTGACCAGGGACAACCGTCA     4679
gi|126033139|gb|EF076724.1|       CAGTGGCCAGGGTAGTTTCTGGGCCAACCTGGGCATAGCCCTGACCAGGGACAACCGTCA     4679
gi|126033149|gb|EF076729.1|       CAGTGGCCAGGGTAGTTTCTGGGCCAACCTGGGCATAGCCCTGACCAGGGACAACCGTCA     4679
gi|126033147|gb|EF076728.1|       CAGTGGCCAGGGTAGTTTCTGGGCCAACCTGGGCATAGCCCTGACCAGGGACAACCGTCA     4679
gi|126033137|gb|EF076723.1|       CAGTGGCCAGGGTAGTTTCTGGGCCAACCTGGGCATAGCCCTGACCAGGGACAACCGTCA     4679
gi|57471257|gb|AY859726.1|        CAGTGGCCAGGGTAGTTTCTGGGCCAACCTGGGCATAGCCCTGACCAGGGACAACCGTCA     3127
gi|27848078|emb|AJ509824.2|       CAGTGGCCAGGGTAGTTTCTGGGCCAACCTGGGCATAGCCCTGACCAGGGACAACCGTCA     4407
gi|34147166|ref|NM_183081.1|      CAGTGGCCAGGGTAGTTTCTGGGCCAACCTGGGCATAGCCCTGACCAGGGACAACCGTCA     3167
gi|23821159|emb|AJ509825.1|       CAGTGGCCAGGGTAGTTTCTGGGCCAACCTGGGCATAGCCCTGACCAGGGACAACCGTCA     3023
                                  ************************************************************

gi|407316560|gb|FJ495080.2|       CTTCTATAACCGGAACTTCTGCCGGGGCCCCACGACAGCCGAATAGCACAGAGTGACTGC     4724
gi|126033143|gb|EF076726.1|       CTTCTATAACCGGAACTTCTGCCGGGGCCCCACGACAGCCGAATAGCACAGAGTGACTGC     4739
gi|126033153|gb|EF076731.1|       CTTCTATAACCGGAACTTCTGCCGGGGCCCCACGACAGCCGAATAGCACAGAGTGACTGC     4739
gi|126033145|gb|EF076727.1|       CTTCTATAACCGGAACTTCTGCCGGGGCCCCACGACAGCCGAATAGCACAGAGTGACTGC     4739
gi|126033141|gb|EF076725.1|       CTTCTATAACCGGAACTTCTGCCGGGGCCCCACGACAGCCGAATAGCACAGAGTGACTGC     4739
gi|126033139|gb|EF076724.1|       CTTCTATAACCGGAACTTCTGCCGGGGCCCCACGACAGCCGAATAGCACAGAGTGACTGC     4739
gi|126033149|gb|EF076729.1|       CTTCTATAACCGGAACTTCTGCCGGGGCCCCACGACAGCCGAATAGCACAGAGTGACTGC     4739
gi|126033147|gb|EF076728.1|       CTTCTATAACCGGAACTTCTGCCGGGGCCCCACGACAGCCGAATAGCACAGAGTGACTGC     4739
gi|126033137|gb|EF076723.1|       CTTCTATAACCGGAACTTCTGCCGGGGCCCCACGACAGCCGAATAGCACAGAGTGACTGC     4739
gi|57471257|gb|AY859726.1|        CTTCTATAACCGGAACTTCTGCCGGGGCCCCACGACAGCCGAATAGCACAGAGTGACTGC     3187
gi|27848078|emb|AJ509824.2|       CTTCTATAACCGGAACTTCTGCCGGGGCCCCACGACAGCCGAATAGCACAGACTGCCTGC     4467
gi|34147166|ref|NM_183081.1|      CTTCTATAACCGGAACTTCTGCCGGGGCCCCACGACAGCCGAATAGCACAGACTGCCTGC     3227
gi|23821159|emb|AJ509825.1|       CTTCTATAACCGGAACTTCTGCCGGGGCCCCACGACAGCCGAATAGCACAGACTGCCTGC     3083
                                  ************************************************** ** ****

gi|407316560|gb|FJ495080.2|       CCAGCACTCCTTTAACCCCACACCCCCACTTCCCACCCCCCCCCCCGCCCGACTCCCCACA     4784
gi|126033143|gb|EF076726.1|       CCAGCACTCCTTTAACCCCACACCCCCACTTCCCACCCCCCCCCCGCCCGACTCCCCACA     4799
gi|126033153|gb|EF076731.1|       CCAGCACTCCTTTAACCCCACACCCCCACTTCCCACCCCCCCCCCGCCCGACTCCCCACA     4799
gi|126033145|gb|EF076727.1|       CCAGCACTCCTTTAACCCCACACCCCCACTTCCCACCCCCCCCCCGCCCGACTCCCCACA     4799
```

```
gi|126033141|gb|EF076725.1|       CCAGCACTCCTTTAACCCCACACCCCCACTTCCCACCCCCCCCCCCGCCCGACTCCCCACA     4799
gi|126033139|gb|EF076724.1|       CCAGCACTCCTTTAACCCCACACCCCCACTTCCCACCCCCCCCCCCGCCCGACTCCCCACA     4799
gi|126033149|gb|EF076729.1|       CCAGCACTCCTTTAACCCCACACCCCCACTTCCCACCCCCCCCCCCGCCCGACTCCCCACA     4799
gi|126033147|gb|EF076728.1|       CCAGCACTCCTTTAACCCCACACCCCCACTTCCCACCCCCCCCCCCGCCCGACTCCCCACA     4799
gi|126033137|gb|EF076723.1|       CCAGCACTCCTTTAACCCCACACCCCCACTTCCCACCCCCCCCCCCGCCCGACTCCCCACA     4799
gi|57471257|gb|AY859726.1|        CCAG-------------------------------------------------------     3191
gi|27848078|emb|AJ509824.2|       CCAGCACTCCTTTAACCCCACACCCCCAGCCCCTAAGA----------------------     4505
gi|34147166|ref|NM_183081.1|      CCAGCACTCCTTTAACCCCACACCCCCAGCCCCTAAGA----------------------     3265
gi|23821159|emb|AJ509825.1|       CCAGCACTCCTTTAACCCCACACCCCCAGCCCCTAAGA----------------------     3121
                                  ****

gi|407316560|gb|FJ495080.2|       ACGCTCCCCCCCCCCCCCGCCCCCACCGTTTTGCCTCTCTGCCTGGGA-------------     4831
gi|126033143|gb|EF076726.1|       ACGCTCCCCCCCCCCCCCGCCCCCACCGTTTTGCCTCTCTGCCTGGGATGCCCCAACTTGC     4859
gi|126033153|gb|EF076731.1|       ACGCTCCCCCCCCCCCCCGCCCCCACCGTTTTGCCTCTCTGCCTGGGATGCCCCAACTTGC     4859
gi|126033145|gb|EF076727.1|       ACGCTCCCCCCCCCCCCCGCCCCCACCGTTTTGCCTCTCTGCCTGGGATGCCCCAACTTGC     4859
gi|126033141|gb|EF076725.1|       ACGCTCCCCCCCCCCCCCGCCCCCACCGTTTTGCCTCTCTGCCTGGGATGCCCCAACTTGC     4859
gi|126033139|gb|EF076724.1|       ACGCTCCCCCCCCCCCCCGCCCCCACCGTTTTGCCTCTCTGCCTGGGATGCCCCAACTTGC     4859
gi|126033149|gb|EF076729.1|       ACGCTCCCCCCCCCCCCCGCCCCCACCGTTTTGCCTCTCTGCCTGGGATGCCCCAACTTGC     4859
gi|126033147|gb|EF076728.1|       ACGCTCCCCCCCCCCCCCGCCCCCACCGTTTTGCCTCTCTGCCTGGGATGCCCCAACTTGC     4859
gi|126033137|gb|EF076723.1|       ACGCTCCCCCCCCCCCCCGCCCCCACCGTTTTGCCTCTCTGCCTGGGATGCCCCAACTTGC     4859
gi|57471257|gb|AY859726.1|        -------------------------------------------------------------     3191
gi|27848078|emb|AJ509824.2|       -------------------------------------------------------------     4505
gi|34147166|ref|NM_183081.1|      -------------------------------------------------------------     3265
gi|23821159|emb|AJ509825.1|       -------------------------------------------------------------     3121


gi|407316560|gb|FJ495080.2|       -------------------------------------------------------------     4831
gi|126033143|gb|EF076726.1|       CTCCCTTTGCAATCCCACCGCTCTGTCTGGCCCCCTGGCATAGAGCAGGCACACAAATAA     4919
gi|126033153|gb|EF076731.1|       CTCCCTTTGCAATCCCACCGCTCTGTCTGGCCCCCTGGCATAGAGCAGGCACACAAATAA     4919
gi|126033145|gb|EF076727.1|       CTCCCTTTGCAATCCCACCGCTCTGTCTGGCCCCCTGGCATAGAGCAGGCACACAAATAA     4919
gi|126033141|gb|EF076725.1|       CTCCCTTTGCAATCCCACCGCTCTGTCTGGCCCCCTGGCATAGAGCAGGCACACAAATAA     4919
gi|126033139|gb|EF076724.1|       CTCCCTTTGCAATCCCACCGCTCTGTCTGGCCCCCTGGCATAGAGCAGGCACACAAATAA     4919
gi|126033149|gb|EF076729.1|       CTCCCTTTGCAATCCCACCGCTCTGTCTGGCCCCCTGGCATAGAGCAGGCACACAAATAA     4919
gi|126033147|gb|EF076728.1|       CTCCCTTTGCAATCCCACCGCTCTGTCTGGCCCCCTGGCATAGAGCAGGCACACAAATAA     4919
gi|126033137|gb|EF076723.1|       CTCCCTTTGCAATCCCACCGCTCTGTCTGGCCCCCTGGCATAGAGCAGGCACACAAATAA     4919
gi|57471257|gb|AY859726.1|        -------------------------------------------------------------     3191
gi|27848078|emb|AJ509824.2|       -------------------------------------------------------------     4505
gi|34147166|ref|NM_183081.1|      -------------------------------------------------------------     3265
gi|23821159|emb|AJ509825.1|       -------------------------------------------------------------     3121
```

```
gi|407316560|gb|FJ495080.2|      ------------------------------------------------------------     4831
gi|126033143|gb|EF076726.1|      ATGCTGCTGGAGGGCCAACAGCTAACCCTGAGCTCACTGAAGGTGCTATTGGGAGGGAAA     4979
gi|126033153|gb|EF076731.1|      ATGCTGCTGGAGGGCCAACAGCTAACCCTGAGCTCACTGAAGGTGCTATTGGGAGGGAAA     4979
gi|126033145|gb|EF076727.1|      ATGCTGCTGGAGGGCCAACAGCTAACCCTGAGCTCACTGAAGGTGCTATTGGGAGGGAAA     4979
gi|126033141|gb|EF076725.1|      ATGCTGCTGGAGGGCCAACAGCTAACCCTGAGCTCACTGAAGGTGCTATTGGGAGGGAAA     4979
gi|126033139|gb|EF076724.1|      ATGCTGCTGGAGGGCCAACAGCTAACCCTGAGCTCACTGAAGGTGCTATTGGGAGGGAAA     4979
gi|126033149|gb|EF076729.1|      ATGCTGCTGGAGGGCCAACAGCTAACCCTGAGCTCACTGAAGGTGCTATTGGGAGGGAAA     4979
gi|126033147|gb|EF076728.1|      ATGCTGCTGGAGGGCCAACAGCTAACCCTGAGCTCACTGAAGGTGCTATTGGGAGGGAAA     4979
gi|126033137|gb|EF076723.1|      ATGCTGCTGGAGGGCCAACAGCTAACCCTGAGCTCACTGAAGGTGCTATTGGGAGGGAAA     4979
gi|57471257|gb|AY859726.1|       ------------------------------------------------------------     3191
gi|27848078|emb|AJ509824.2|      ------------------------------------------------------------     4505
gi|34147166|ref|NM_183081.1|    ------------------------------------------------------------     3265
gi|23821159|emb|AJ509825.1|      ------------------------------------------------------------     3121


gi|407316560|gb|FJ495080.2|      ------------------------------------------------------    4831
gi|126033143|gb|EF076726.1|      AGGGGAGAATTCACTGGGCCTAAAACAAAGGCGGGGAGCAAGAGGTGATTGGCG    5033
gi|126033153|gb|EF076731.1|      AGGGGAGAATTCACTGGGCCTAAAACAAAGGCGGGGAGCAAGAGGTGATTGGCG    5033
gi|126033145|gb|EF076727.1|      AGGGGAGAATTCACTGGGCCTAAAACAAAGGCGGGGAGCAAGAGGTGATTGGCG    5033
gi|126033141|gb|EF076725.1|      AGGGGAGAATTCACTGGGCCTAAAACAAAGGCGGGGAGCAAGAGGTGATTGGCG    5033
gi|126033139|gb|EF076724.1|      AGGGGAGAATTCACTGGGCCTAAAACAAAGGCGGGGAGCAAGAGGTGATTGGCG    5033
gi|126033149|gb|EF076729.1|      AGGGGAGAATTCACTGGGCCTAAAACAAAGGCGGGGAGCAAGAGGTGATTGGCG    5033
gi|126033147|gb|EF076728.1|      AGGGGAGAATTCACTGGGCCTAAAACAAAGGCGGGGAGCAAGAGGTGATTGGCG    5033
gi|126033137|gb|EF076723.1|      AGGGGAGAATTCACTGGGCCTAAAACAAAGGCGGGGAGCAAGAGGTGATTGGCG    5033
gi|57471257|gb|AY859726.1|       ------------------------------------------------------    3191
gi|27848078|emb|AJ509824.2|      ------------------------------------------------------    4505
gi|34147166|ref|NM_183081.1|    ------------------------------------------------------    3265
gi|23821159|emb|AJ509825.1|      ------------------------------------------------------    3121
```

Appendix H: TLR9 Polymorphism data

Sample ID represents the code name for bovine samples. SNP allele represents the polymorphisms that can be found at the locus. db SNP allele represents what allele for SNP reported in NCBI SNP databse. Nucleotide position represent the locus of TLR9 were polymorphism is found. Heterozygosity represents the genotype. SNP on peaks represents the visualisation of SNP in chromatogram. Novelty represents if SNP has been previously reported. aa represents amino acid change, aa position represents amino acid position, CpG Island represent if SNP is found in CpG Island.

Sample ID	Polymorphism		Nucleotide position	Heterozygosity	SNP on Peaks	Novelty	aa change	aa position	CpG Island
	SNP Allele	db SNP allele							
T1	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Heterozygous	Yes	No	-----	----	-----
	A/G	A/G	2583	Het	Yes	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
T2	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Heterozygous	Yes	No	-----	----	-----
	A/G	A/G	2583	Het	Yes	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
T3	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Homozygous	No	No	-----	----	-----
	A/G	A/G	2547	Het	Yes	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
T5	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Homozygous	No	No	-----	----	-----
	A/G	A/G	2547	Het	Yes	No	-----	----	-----

	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
T6	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Homo	No	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
T7	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	A/G	A/G	2583	Homo	No	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
T8	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2256	Heterozygous	Yes	No	-----	----	Island 1
	C/T	C/T	2538	Heterozygous	Yes	No	-----	----	-----
	A/G	A/G	2583	Het	Yes	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
T9	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Homo	No	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
T10	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Homozygous	No	No	-----	----	-----
	A/G	A/G	2547	Het	Yes	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
T11	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Homo	No	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
T12	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
T13	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1

	C/T	C/T	2538	Homo	No	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
T14	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Homozygous	No	No	-----	----	-----
	A/G	A/G	2547	Hete	Yes	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
T15	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	A/G	A/G	2583	Homo	No	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
T16	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2256	Homo	No	No	----	----	Island 1
	C/T	C/T	2538	Homo	No	No	-----	----	-----
	A/G	A/G	2583	Homo	No	No	-----	-----	-----
	G/C	G/C	2823	Homo	No	No	-----	-----	Island 2
	A/G	A/G	2865	Homo	No	No	-----	------	Island 2
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
T17	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Hete	Yes	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
T18	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Heterozygous	Yes	No	-----	----	-----
	A/G	A/G	2583	Het	Yes	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
T19	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2256	Hete	Yes	No	----	----	Island 1
	C/T	C/T	2538	Homo	No	No	-----	----	-----
	G/C	G/C	2823	Hete	Yes	No	-----	-----	Island 2

	A/G	A/G	2865	Hete	Yes	No	-----	------	Island 2
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
T20	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Homo	No	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
B1	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Heterozygous	Yes	No	-----	----	-----
	A/G	A/G	2583	Het	Yes	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
B2	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Homo	No	No	-----	----	-----
	A/G	A/G	2583	Het	Yes	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
B3	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2256	Hete	Yes	No	----	----	Island 1
	C/T	C/T	2538	Hete	Yes	No	-----	----	-----
	A/G	A/G	2583	Hete	Yes	No	-----	-----	-----
	G/C	G/C	2823	Hete	Yes	No	-----	-----	Island 2
	A/G	A/G	2865	Hete	Yes	No	-----	------	Island 2
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
B4	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Hete	Yes	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
B5	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2256	Homo	No	No	----	----	Island 1
	C/T	C/T	2538	Homo	No	No	-----	----	-----
	A/G	A/G	2583	Homo	No	No	-----	-----	-----

	G/C	G/C	2823	Homo	No	No	-----	-----	Island 2
	A/G	A/G	2865	Homo	No	No	-----	------	Island 2
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
B6	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2256	Hete	Yes	No	----	----	Island 1
	C/T	C/T	2538	Hete	Yes	No	-----	----	-----
	G/C	G/C	2823	Hete	Yes	No	-----	-----	Island 2
	A/G	A/G	2865	Homo	No	No	-----	------	Island 2
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
B7	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Homozygous	No	No	-----	----	-----
	A/G	A/G	2547	Het	Yes	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
B8	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	A/G	A/G	2583	Homo	No	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
B9	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2256	Hete	Yes	No	----	----	Island 1
	C/T	C/T	2538	Homo	No	No	-----	----	-----
	G/C	G/C	2823	Hete	Yes	No	-----	-----	Island 2
	A/G	A/G	2865	Hete	Yes	No	-----	------	Island 2
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
B10	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2256	Hete	Yes	No	----	----	Island 1
	C/T	C/T	2538	Hete	Yes	No	-----	----	-----
	A/G	A/G	2583	Hete	Yes	No	-----	-----	-----
	G/C	G/C	2823	Hete	Yes	No	-----	-----	Island 2
	A/G	A/G	2865	Hete	Yes	No	-----	------	Island 2

	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
B11	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Homo	No	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
B12	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2256	Hete	Yes	No	----	----	Island 1
	C/T	C/T	2538	Hete	Yes	No	-----	----	-----
	G/C	G/C	2823	Hete	Yes	No	-----	-----	Island 2
	A/G	A/G	2865	Homo	No	No	-----	------	Island 2
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
B14	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Heterozygous	Yes	No	-----	----	-----
	A/G	A/G	2583	Het	Yes	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
B15	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Hete	Yes	No	-----	----	-----
	A/G	A/G	2547	Het	Yes	No	-----	----	-----
	A/G	A/G	2583	Het	Yes	No	-----	-----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
B16	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Heterozygous	Yes	No	-----	----	-----
	A/G	A/G	2583	Het	Yes	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
B17	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	A/G	A/G	2583	Homo	No	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2

B18	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Homo	No	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
B19	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Hete	Yes	No	-----	----	-----
	A/G	A/G	2547	Het	Yes	No	-----	----	-----
	A/G	A/G	2583	Het	Yes	No	-----	-----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
B20	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Homo	No	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
K1	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Heterozygous	Yes	No	-----	----	-----
	A/G	A/G	2583	Het	Yes	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
K2	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2256	Homo	No	No	----	----	Island 1
	C/T	C/T	2538	Homo	No	No	-----	----	-----
	A/G	A/G	2583	Homo	No	No	-----	-----	-----
	G/C	G/C	2823	Homo	No	No	-----	-----	Island 2
	A/G	A/G	2865	Homo	No	No	-----	------	Island 2
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
K3	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Homo	No	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
K4	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Hete	Yes	No	-----	----	-----

	A/G	A/G	2547	Het	Yes	No	-----	----	-----
	A/G	A/G	2583	Het	Yes	No	-----	-----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
K5	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Hete	Yes	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
K6	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Hete	Yes	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
K7	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Homozygous	No	No	-----	----	-----
	A/G	A/G	2547	Homo	No	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
K8	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Homo	No	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
K9	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Heterozygous	Yes	No	-----	----	-----
	A/G	A/G	2583	Het	Yes	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
K10	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Homo	No	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
K11	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Homo	No	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2

	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
K12	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2256	Hete	Yes	No	----	----	-----
	C/T	C/T	2538	Homo	Yes	No	-----	-----	-----
	G/C	G/C	2823	Hete	Yes	No	-----	-----	------
	A/G	A/G	2865	Hete	Yes	No	-----	------	-------
	C/T	C/T	2916	Homo	No	Yes	-----	-----	-----
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
K13	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Homo	No	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
K14	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2256	Hete	Yes	No	----	----	Island 1
	C/T	C/T	2538	Homo	No	No	-----	----	-----
	A/G	A/G	2547	Hete	Yes	No	-----	-----	-----
	G/C	G/C	2823	Hete	Yes	No	-----	-----	Island 2
	A/G	A/G	2865	Hete	Yes	No	-----	------	Island 2
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
K15	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2256	Hete	Yes	No	----	----	-----
	C/T	C/T	2538	Homo	Yes	No	-----	-----	-----
	G/C	G/C	2823	Hete	Yes	No	-----	-----	------
	A/G	A/G	2865	Hete	Yes	No	-----	------	-------
	C/T	C/T	2916	Homo	No	Yes	-----	-----	-----
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
K16	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Homozygous	No	No	-----	----	-----
	A/G	A/G	2547	Hete	Yes	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
K17	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1

	C/T	C/T	2256	Hete	Yes	No	----	----	-----
	C/T	C/T	2538	Homo	Yes	No	-----	-----	-----
	G/C	G/C	2823	Hete	Yes	No	-----	-----	------
	A/G	A/G	2865	Hete	Yes	No	-----	------	-------
	C/T	C/T	2916	Homo	No	Yes	-----	-----	-----
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
K18	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Hete	Yes	No	-----	----	-----
	A/G	A/G	2547	Het	Yes	No	-----	----	-----
	A/G	A/G	2583	Het	Yes	No	-----	-----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
K19	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2256	Hete	Yes	No	----	----	-----
	C/T	C/T	2538	Homo	Yes	No	-----	-----	-----
	G/C	G/C	2823	Hete	Yes	No	-----	-----	------
	A/G	A/G	2865	Hete	Yes	No	-----	------	-------
	C/T	C/T	2916	Homo	No	Yes	-----	-----	-----
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
U5	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Homo	No	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
U6	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2256	Hete	Yes	No	----	----	Island 1
	C/T	C/T	2538	Hete	Yes	No	-----	----	-----
	A/G	A/G	2583	Hete	Yes	No	-----	-----	-----
	G/C	G/C	2823	Hete	Yes	No	-----	-----	Island 2
	A/G	A/G	2865	Hete	Yes	No	-----	------	Island 2
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
U7	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Heterozygous	Yes	No	-----	----	-----

	A/G	A/G	2583	Het	Yes	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
U8	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2256	Hete	Yes	No	----	----	-----
	C/T	C/T	2538	Hete	Yes	No	-----	-----	-----
	G/C	G/C	2823	Hete	Yes	No	-----	-----	------
	A/G	A/G	2865	Hete	Yes	No	-----	------	-------
	C/T	C/T	2916	Homo	No	Yes	-----	-----	-----
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
U9	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	A/G	A/G	2583	Homo	No	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
U10									
U11	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2256	Hete	Yes	No	----	----	Island 1
	C/T	C/T	2538	Hete	Yes	No	-----	----	-----
	A/G	A/G	2583	Hete	Yes	No	-----	-----	-----
	G/C	G/C	2823	Hete	Yes	No	-----	-----	Island 2
	A/G	A/G	2865	Hete	Yes	No	-----	------	Island 2
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
U12	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	A/G	A/G	2583	Homo	No	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
U61	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	A/G	A/G	2583	Homo	No	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
U62	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Homo	No	No	-----	----	-----

	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
U63	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	A/G	A/G	2583	Homo	No	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
U64	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2256	Hete	Yes	No	----	----	-----
	C/T	C/T	2538	Homo	Yes	No	-----	-----	-----
	G/C	G/C	2823	Hete	Yes	No	-----	-----	------
	A/G	A/G	2865	Hete	Yes	No	-----	------	-------
	C/T	C/T	2916	Homo	No	Yes	-----	-----	-----
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
U85	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2256	Hete	Yes	No	----	----	-----
	C/T	C/T	2538	Homo	Yes	No	-----	-----	-----
	G/C	G/C	2823	Hete	Yes	No	-----	-----	Island 2
	A/G	A/G	2865	Hete	Yes	No	-----	------	Island 2
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
U86	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Homo	No	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
U87	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2538	Homo	No	No	-----	----	-----
	C/T	C/T	2916	Homo	No	Yes	-----	-----	Island 2
	A/C	A/C	2924	Homo	No	Yes	D/A	934	Island 2
U88	C/G	C/G	2238	Homozygous	No	Yes	D/E	705	Island 1
	C/T	C/T	2256	Homo	No	No	----	----	-----
	C/T	C/T	2538	Homo	No	No	-----	-----	-----
	G/C	G/C	2823	Homo	No	No	-----	-----	------
	A/G	A/G	2865	Homo	No	No	-----	------	-------

	C/T A/C	C/T A/C	2916 2924	Homo Homo	No No	Yes Yes	----- D/A	----- 934	----- Island 2
U93	C/G C/T C/T G/C A/G C/T A/C	C/G C/T C/T G/C A/G C/T A/C	2238 2256 2538 2823 2865 2916 2924	Homozygous Hete Homo Hete Hete Homo Homo	No Yes Yes Yes Yes No No	Yes No No No No Yes Yes	D/E ---- ----- ----- ----- ----- D/A	705 ---- ----- ----- ------ ----- 934	Island 1 ----- ----- ------ ------- ----- Island 2

References

Abenga, JN, Enwezor, FNC, Lawani, FAG, Osue, HO, & Ikemereh, ECD. (2004). Trypanosome prevalence in cattle in Lere area in Kaduna State, North Central Nigeria. *Revue d'élevage et de médecine vétérinaire des pays tropicaux, 57*(1-2), 45-48.

Abreu, M. T., & Arditi, M. (2004). Innate immunity and toll-like receptors: clinical implications of basic science research. *The Journal of Pediatrics, 144*(4), 421-429. doi: 10.1016/j.jpeds.2004.01.057

Acosta Ida, C., da Costa, A. P., Nunes, P. H., Gondim, M. F., Gatti, A., Rossi, J. L., Jr., . . . Marcili, A. (2013). Morphological and molecular characterization and phylogenetic relationships of a new species of trypanosome in Tapirus terrestris (lowland tapir), *Trypanosoma* terrestris sp. nov., from Atlantic Rainforest of southeastern Brazi. *Parasites & vectors, 6*, 349. doi: 10.1186/1756-3305-6-349

Adams, ER, Hamilton, PB, Malele, II, & Gibson, WC. (2008). The identification, diversity and prevalence of trypanosomes in field caught tsetse in Tanzania using ITS-1 primers and fluorescent fragment length barcoding. *Infection, Genetics and Evolution, 8*(4), 439-444.

Ademosun, AA. (1973). The development of the livestock industry in Nigeria-Ruminants. *Proceedings of the Agricultural Society of Niger, 10*, 13-20.

Ahmad, S., Khan, Z., Mustafa, A. S., & Khan, Z. U. (2002). Seminested PCR for diagnosis of candidemia: comparison with culture, antigen detection, and biochemical methods for species identification. *Journal of clinical microbiology, 40*(7), 2483-2489.

Ahmed, H. A., MacLeod, E. T., Hide, G., Welburn, S. C., & Picozzi, K. (2011). The best practice for preparation of samples from FTA(R)cards for diagnosis of blood borne infections using African trypanosomes as a model system. *Parasites & vectors, 4*, 68. doi: 10.1186/1756-3305-4-68

Ahmed, H. A., Picozzi, K., Welburn, S. C., & MacLeod, E. T. (2013). A comparative evaluation of PCR-based methods for species- specific determination of African animal trypanosomes in Ugandan cattle. *Parasites & vectors, 6*(1), 316. doi: 10.1186/1756-3305-6-316

Airauhi, LU, Airauhi, EA, & Adesina, KA. (2008). Sleeping Sickness Surveillance In The Abraka Sleeping Sickness Focus, Nigeria. *Annals of Biomedical Sciences, 2*(1), 20-29.

Akira, S. (2003). Mammalian Toll-like receptors. *Current opinion in immunology, 15*(1), 5-11.

Akira, Shizuo, & Takeda, Kiyoshi. (2004). Toll-like receptor signalling. *Nature reviews immunology, 4*(7), 499-511.

Aksoy, S. (2003). Control of tsetse flies and trypanosomes using molecular genetics. *Veterinary parasitology, 115*(2), 125-145.

Anene, B. M., Chime, A. B., Jibike, G. I., & Anika, S. M. (1991). Comparative study of clinical signs, haematology and prevalence of trypanosomiasis in Holstein Friesian and White Fulani Zebu cattle exposed to natural infection in a rain forest zone of Nigeria. *Angewandte Parasitologie, 32*(2), 99-104.

Arai, E., Kanai, Y., Ushijima, S., Fujimoto, H., Mukai, K., & Hirohashi, S. (2006). Regional DNA hypermethylation and DNA methyltransferase (DNMT) 1 protein overexpression in both renal tumors and corresponding nontumorous renal tissues. *International journal of cancer. Journal international du cancer, 119*(2), 288-296. doi: 10.1002/ijc.21807

Bachman, K. E., Argani, P., Samuels, Y., Silliman, N., Ptak, J., Szabo, S., . . . Park, B. H. (2004). The PIK3CA gene is mutated with high frequency in human breast cancers. *Cancer biology & therapy, 3*(8), 772-775.

Bajnok, J., Boyce, K., Rogan, M. T., Craig, P. S., Lun, Z. R., & Hide, G. (2014). Prevalence of Toxoplasma gondii in localized populations of Apodemus sylvaticus is linked to population genotype not to population location. *Parasitology*, 1-11. doi: 10.1017/s0031182014001760

Bardsley, J E, & Harmsen, R. (1973). The trypanosomes of anura. *Advances in Parasitology, 11*(0), 1-73.

Barrett, M. P., Burchmore, R. J., Stich, A., Lazzari, J. O., Frasch, A. C., Cazzulo, J. J., & Krishna, S. (2003). The trypanosomiases. *Lancet, 362*(9394), 1469-1480. doi: 10.1016/s0140-6736(03)14694-6

Bolboaca, Sorana-Daniela, & Jäntschi, Lorentz. (2006). Pearson versus Spearman, Kendall's tau correlation analysis on structure-activity relationships of biologic active compounds. *Leonardo Journal of Sciences, 5*(9), 179-200.

Bose, R., & Heister, N. C. (1993). Development of *Trypanosoma (M.) theileri* in tabanids. *The Journal of eukaryotic microbiology, 40*(6), 788-792.

Bromberg, Y., & Rost, B. (2007). SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Research, 35*(11), 3823-3835. doi: 10.1093/nar/gkm238

Brun, R., Blum, J., Chappuis, F., & Burri, C. (2010). Human African trypanosomiasis. *Lancet, 375*(9709), 148-159. doi: 10.1016/s0140-6736(09)60829-1

Bugaje, IM. (2006). Renewable energy for sustainable development in Africa: a review. *Renewable and Sustainable Energy Reviews, 10*(6), 603-612.

Büscher, Philippe. (2014). Diagnosis of African Trypanosomiasis *Trypanosomes and Trypanosomiasis* (pp. 189-216): Springer.

Cai, T. T., Muhali, F. S., Song, R. H., Qin, Q., Wang, X., Shi, L. F., . . . Zhang, J. A. (2015). Genome-wide DNA methylation analysis in Graves' disease. *Genomics, 105*(4), 204-210. doi: 10.1016/j.ygeno.2015.01.001

Campos, M. A., Almeida, I. C., Takeuchi, O., Akira, S., Valente, E. P., Procopio, D. O., . . . Gazzinelli, R. T. (2001). Activation of Toll-like receptor-2 by glycosylphosphatidylinositol anchors from a protozoan parasite. *Journal of immunology, 167*(1), 416-423.

Capewell, P., Veitch, N. J., Turner, C. M., Raper, J., Berriman, M., Hajduk, S. L., & MacLeod, A. (2011). Differences between *Trypanosoma* brucei gambiense groups 1 and 2 in their resistance to killing by trypanolytic factor 1. *PLoS neglected tropical diseases, 5*(9), e1287. doi: 10.1371/journal.pntd.0001287

Cargill, E. J., & Womack, J. E. (2007). Detection of polymorphisms in bovine toll-like receptors 3, 7, 8, and 9. *Genomics, 89*(6), 745-755. doi: 10.1016/j.ygeno.2007.02.008

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution, 17*(4), 540-552.

Chanie, Mersha, Adula, Dulecha, & Bogale, Basaznew. (2013). Socio-Economic Assessment of the Impacts of Trypanosomiasis on Cattle in Girja District, Southern Oromia Region, Southern Ethiopia. *Acta Parasitologica Globalis, 4*, 80-85.

Chappuis, F., Loutan, L., Simarro, P., Lejon, V., & Buscher, P. (2005). Options for field diagnosis of human african trypanosomiasis. *Clinical microbiology reviews, 18*(1), 133-146. doi: 10.1128/cmr.18.1.133-146.2005

Cnops, L., Tannich, E., Polman, K., Clerinx, J., & Van Esbroeck, M. (2012). Schistosoma real-time PCR as diagnostic tool for international travellers and migrants. *Tropical medicine & international health, 17*(10), 1208-1216. doi: 10.1111/j.1365-3156.2012.03060.x

Corner, L.A.L, Murphy, D, & Gormley, E. (2011). Mycobacterium bovis Infection in the Eurasian Badger (Meles meles): the Disease, Pathogenesis, Epidemiology and Control. *Journal of Comparative Pathology, 144*, 1-24.

Costello, Joseph F, & Plass, Christoph. (2001). Methylation matters. *Journal of Medical Genetics, 38*(5), 285-303.

Cottontail, V. M., Kalko, E. K., Cottontail, I., Wellinghausen, N., Tschapka, M., Perkins, S. L., & Pinto, C. M. (2014). High local diversity of *Trypanosoma* in a common bat species, and implications for the biogeography and taxonomy of the T. cruzi clade. *PLoS One, 9*(9), e108603. doi: 10.1371/journal.pone.0108603

Courtin, D., Berthier, D., Thevenon, S., Dayo, G. K., Garcia, A., & Bucheton, B. (2008). Host genetics in African trypanosomiasis. *Infection Genetics Evolution, 8*(3), 229-238. doi: 10.1016/j.meegid.2008.02.007

Cox, A, Tilley, A, McOdimba, F, Fyfe, J, Hide, G, & Welburn, S.C. (2005). A PCR based assay for detection and differentiation of African trypanosome species in blood. *Experimental Parasitology, 111*, 24-29.

Cox, A. P., Tosas, O., Tilley, A., Picozzi, K., Coleman, P., Hide, G., & Welburn, S. C. (2010). Constraints to estimating the prevalence of trypanosome infections in East African zebu cattle. *Parasites & vectors 6*(3), 82. doi: 10.1186/1756-3305-3-82

Cox, D. R., Burmeister, M., Price, E. R., Kim, S., & Myers, R. M. (1990). Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science, 250*(4978), 245-250.

Da Silva, F M, Noyes, H, Campaner, M, Junqueira, A C, Coura, J R, Añez, N, . . . Teixeira, M M. (2004). Phylogeny, taxonomy and grouping of *Trypanosoma rangeli* isolates from man, triatomines and sylvatic mammals from widespread geographical origin based on SSU and ITS ribosomal sequences. *Parasitology, 129*(Pt5), 549-561.

Daley, D. (2014). The evolution of the hygiene hypothesis: the role of early-life exposures to viruses and microbes and their relationship to asthma and allergic diseases. *Current opinion in allergy and clinical immunology, 14*(5), 390-396. doi: 10.1097/aci.0000000000000101

Daniel, A. D., Dadah, A. J., Kalejaiye, J. O., & Dalhatu, A. D. (1993). Prevalence of bovine trypanosomiasis in Gongola State of Northern Nigeria. *Revue d'elevage et de medecine veterinaire des pays tropicaux, 46*(4), 571-574.

Dayeh, T. A., Olsson, A. H., Volkov, P., Almgren, P., Ronn, T., & Ling, C. (2013). Identification of CpG-SNPs associated with type 2 diabetes and differential DNA methylation in human pancreatic islets. *Diabetologia, 56*(5), 1036-1046. doi: 10.1007/s00125-012-2815-7

De Jager, P. L., Srivastava, G., Lunnon, K., Burgess, J., & Schalkwyk, L. C. (2014). Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nature Neuroscience, 17*(9), 1156-1163. doi: 10.1038/nn.3786

Deborggraeve, S., & Buscher, P. (2010). Molecular diagnostics for sleeping sickness: what is the benefit for the patient? *The Lancet Infectious Diseases, 10*(6), 433-439. doi: 10.1016/s1473-3099(10)70077-3

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., . . . Gascuel, O. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research, 36*(Web Server issue), W465-469. doi: 10.1093/nar/gkn180

Desquesnes, M., Kamyingkird, K., Yangtara, S., Milocco, C., Ravel, S., Wang, M. H., . . . Jittapalapong, S. (2011). Specific primers for PCR amplification of the ITS1 (ribosomal DNA) of *Trypanosoma lewisi*. *Infection, Genetics and Evolution, 11*(6), 1361-1367. doi: 10.1016/j.meegid.2011.04.030

Djordjevic, SP, Noone, K, Smith, L, & Hornitzky, MAZ. (1998). Development of a hemi-nested PCR assay for the specific detection of Melissococcus pluton. *Journal of Apicultural Research, 37*(3), 165-174.

Drennan, M. B., Stijlemans, B., Van den Abbeele, J., Quesniaux, V. J., Barkhuizen, M., Brombacher, F., . . . Magez, S. (2005). The induction of a type 1 immune response following a *Trypanosoma brucei* infection is MyD88 dependent. *Journal of Immunology 175*(4), 2501-2509.

Dukes, P., Gibson, W. C., Gashumba, J. K., Hudson, K. M., Bromidge, T. J., Kaukus, A., . . . Magnus, E. (1992). Absence of the LiTat 1.3 (CATT antigen) gene in Trypanosoma brucei gambiense stocks from Cameroon. *Acta Trop, 51*(2), 123-134.

Duncanson, P., Terry, R. S., Smith, J. E., & Hide, G. (2001). High levels of congenital transmission of *Toxoplasma gondii* in a commercial sheep flock. *International Journal for Parasitology, 31*(14), 1699-1703.

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics, 5*, 113. doi: 10.1186/1471-2105-5-113

Ehrlich, M., Gama-Sosa, M. A., Huang, L. H., Midgett, R. M., Kuo, K. C., McCune, R. A., & Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Research, 10*(8), 2709-2721.

Enanga, B., Burchmore, R. J., Stewart, M. L., & Barrett, M. P. (2002). Sleeping sickness and the brain. *Cellular and Molecular Life Sciences, 59*(5), 845-858.

Enwezor, F. N., Samdi, S. M., Ijabor, O., & Abenga, J. N. (2012). The prevalence of bovine trypanosomes in parts of Benue state, north-central Nigeria. *Journal of Vector Borne Diseases, 49*(3), 188-190.

Enwezor, F. N., Umoh, J. U., Esievo, K. A., Halid, I., Zaria, L. T., & Anere, J. I. (2009). Survey of bovine trypanosomosis in the Kachia Grazing Reserve, Kaduna State, Nigeria. *Veterinary Parasitology, 159*(2), 121-125. doi: 10.1016/j.vetpar.2008.10.032

Enyaru, J. C., Matovu, E., Akol, M., Sebikali, C., Kyambadde, J., Schmidt, C., . . . Kansiime, F. (1998). Parasitological detection of *Trypanosoma brucei gambiense* in serologically negative sleeping-sickness suspects from north-western Uganda. *Annals of Tropical Medicine and Parasitology, 92*(8), 845-850.

Enyaru, J. C., Matovu, E., Nerima, B., Akol, M., & Sebikali, C. (2006). Detection of *T.b. rhodesiense* trypanosomes in humans and domestic animals in south east Uganda by amplification of serum resistance-associated gene. *Annals of the New York Academy of Sciences, 1081*, 311-319. doi: 10.1196/annals.1373.041

Eperon, G., Balasegaram, M., Potet, J., Mowbray, C., Valverde, O., & Chappuis, F. (2014). Treatment options for second-stage gambiense human African trypanosomiasis. *Expert Review of Anti-infective Therapy, 12*(11), 1407-1417. doi: 10.1586/14787210.2014.959496

Gadagkar, S. R., Rosenberg, M. S., & Kumar, S. (2005). Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. *Journal of experimental zoology. Part B, Molecular and Developmental Evolution, 304*(1), 64-74. doi: 10.1002/jez.b.21026

Gazzinelli, R. T., & Denkers, E. Y. (2006). Protozoan encounters with Toll-like receptor signalling pathways: implications for host parasitism. *Nature Reviews Immunology, 6*(12), 895-906. doi: 10.1038/nri1978

Goldammer, T., Zerbe, H., Molenaar, A., Schuberth, H. J., Brunner, R. M., Kata, S. R., & Seyfert, H. M. (2004). Mastitis increases mammary mRNA abundance of beta-defensin 5, toll-like-receptor 2 (TLR2), and TLR4 but not TLR9 in cattle. *Clinical and Diagnostic Laboratory Immunology, 11*(1), 174-185.

Griebel, P. J., Brownlie, R., Manuja, A., Nichani, A., Mookherjee, N., Popowych, Y., . . . Babiuk, L. A. (2005). Bovine toll-like receptor 9: a comparative analysis of molecular structure, function and expression. *Veterinary Immunology and Immunopathology, 108*(1-2), 11-16. doi: 10.1016/j.vetimm.2005.07.012

Gu, Z, Wang, J, Ke, X, Liu, Y, Liu, X, Gong, X, & Li, A. (2010). Phylogenetic position of the freshwater fish trypanosome, *Trypanosoma ophiocephali (Kinetoplastida)* inferred from the complete small subunit ribosomal RNA gene sequence. *Parasitology Research, 106*(5), 1039-1042.

Haag, J., O'HUigin, C., & Overath, P. (1998). The molecular phylogeny of trypanosomes: evidence for an early divergence of the Salivaria. *Molecular and Biochemical Parasitology, 91*(1), 37-49.

Hahn, P. F., Eisenberg, P. J., Pitman, M. B., Gazelle, G. S., & Mueller, P. R. (1995). Cytopathologic touch preparations (imprints) from core needle biopsies: accuracy compared with that of fine-needle aspirates. *American Journal of Roentgenology, 165*(5), 1277-1279. doi: 10.2214/ajr.165.5.7572518

Hamilton, P. B., Stevens, J. R., Gaunt, M. W., Gidley, J., & Gibson, W. C. (2004). Trypanosomes are monophyletic: evidence from genes for glyceraldehyde phosphate dehydrogenase and small

subunit ribosomal RNA. *International Journal for Parasitology, 34*(12), 1393-1404. doi: 10.1016/j.ijpara.2004.08.011
Haore, C A. (1972). *Hoare, C.A. (1972). The trypanosomes of mammals. Oxford: Blackwell Scientific Publication*. Oxford: Blackwell Scientific Publication.
Harris, T. H., Mansfield, J. M., & Paulnock, D. M. (2007). CpG oligodeoxynucleotide treatment enhances innate resistance and acquired immunity to African trypanosomes. *Infection and Immunity, 75*(5), 2366-2373. doi: 10.1128/iai.01649-06
Hedges, S. B., Moberg, K. D., & Maxson, L. R. (1990). Tetrapod phylogeny inferred from 18S and 28S ribosomal RNA sequences and a review of the evidence for amniote relationships. *Molecular Biology and Evolution, 7*(6), 607-633.
Heller, G., Babinsky, V. N., Ziegler, B., Weinzierl, M., Noll, C., Altenberger, C., . . . Zochbauer-Muller, S. (2013). Genome-wide CpG island methylation analyses in non-small cell lung cancer patients. *Carcinogenesis, 34*(3), 513-521. doi: 10.1093/carcin/bgs363
Hemmi, H., Takeuchi, O., Kawai, T., Kaisho, T., Sato, S., Sanjo, H., . . . Akira, S. (2000). A Toll-like receptor recognizes bacterial DNA. *Nature, 408*(6813), 740-745. doi: 10.1038/35047123
Hernandez, H. G., Tse, M. Y., Pang, S. C., Arboleda, H., & Forero, D. A. (2013). Optimizing methodologies for PCR-based DNA methylation analysis. *Biotechniques, 55*(4), 181-197. doi: 10.2144/000114087
Hide, G. (2003). The origins of sleeping sickness epidemics. *Biologist, 50*(6), 268.
Hide, G, & Tait, A. (2009). Molecular epidemiology of African sleeping sick-ness. *Parasitology, 136*, 1491-1500.
Hide, G., Tait, A., Maudlin, I., & Welburn, S. C. (1996). The origins, dynamics and generation of *Trypanosoma brucei rhodesiense* epidemics in East Africa. *Parasitology Today, 12*(2), 50-55.
Hide, Geoff. (1998). The molecular epidemiology of parasites. *Principles of Medical Biology, 9*, 597-614.
Hutchinson, O. C., Fevre, E. M., Carrington, M., & Welburn, S. C. (2003). Lessons learned from the emergence of a new *Trypanosoma brucei rhodesiense* sleeping sickness focus in Uganda. *The Lancet Infectious diseases, 3*(1), 42-45.
Ikede, BO, Reynolds, L, Ogunsanmi, AO, Fawumi, MK, Ekwuruke, JO, & Taiwo, VO. (1988). The epizootiology of bovine trypanosomiasis in the derived Savannah zone of Nigeria-a preliminary report. *Publication International Scientific Council for Trypanosomiasis Research and Control*(114), 289.
Illingworth, R. S., & Bird, A. P. (2009). CpG islands--'a rough guide'. *FEBS Letters, 583*(11), 1713-1720. doi: 10.1016/j.febslet.2009.04.012
Jamonneau, Vincent, Ilboudo, Hamidou, Kaboré, Jacques, Kaba, Dramane, Koffi, Mathurin, Solano, Philippe, . . . Lingue, Kouakou. (2012). Untreated human infections by T*rypanosoma brucei gambiense* are not 100% fatal. *PLoS Neglected Tropical Diseases, 6*(6), e1691.
Jones, P. A., & Laird, P. W. (1999). Cancer epigenetics comes of age. *Nature Genetics, 21*(2), 163-167. doi: 10.1038/5947
Joshi, P. P., Shegokar, V. R., Powar, R. M., Herder, S., Katti, R., Salkar, H. R., . . . Truc, P. (2005). Human trypanosomiasis caused by *Trypanosoma evansi* in India: the first case report. *The American Journal of Tropical Medicine and Hygiene, 73*(3), 491-495.
Kalu, Arunsi Ukairo. (1995). Prevalence of trypanosomiasis among trypanotolerant cattle at the lower Benue river area of Nigeria. *Preventive Veterinary Medicine, 24*(2), 97-103.
Kanhaiya, Agrawal NK, & Gupta, SK. (2013). Differential expression of toll like receptor 4 in type 2 diabetic patients with impaired wound healing. *Journal of Diabetes & Metabolism*.
Kennedy, P. G. (2004). Human African trypanosomiasis of the CNS: current issues and challenges. *The Journal of Clinical Investigation, 113*(4), 496-504. doi: 10.1172/jci21052
Kennedy, P. G. (2013). Clinical features, diagnosis, and treatment of human African trypanosomiasis (sleeping sickness). *The Lancet Neurology, 12*(2), 186-194. doi: 10.1016/s1474-4422(12)70296-x

Khamechian, T., Alizargar, J., & Mazoochi, T. (2012). The value of touch preparation for rapid diagnosis of brain tumors as an intraoperative consultation. *Iranian Journal of Medical Sciences, 37*(2), 105-111.

Khulan, B., Thompson, R. F., Ye, K., Fazzari, M. J., Suzuki, M., Stasiek, E., . . . Greally, J. M. (2006). Comparative isoschizomer profiling of cytosine methylation: the HELP assay. *Genome Research, 16*(8), 1046-1055. doi: 10.1101/gr.5273806

Kieft, R., Capewell, P., Turner, C. M., Veitch, N. J., MacLeod, A., & Hajduk, S. (2010). Mechanism of *Trypanosoma brucei gambiense* (group 1) resistance to human trypanosome lytic factor. *Proceedings of the National Academy of Sciences of the United States of America, 107*(37), 16137-16141. doi: 10.1073/pnas.1007074107

Kimura, M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution, 16*, 111-120.

Kitamura, E., Igarashi, J., Morohashi, A., Hida, N., Oinuma, T., Nemoto, N., . . . Nagase, H. (2007). Analysis of tissue-specific differentially methylated regions (TDMs) in humans. *Genomics, 89*(3), 326-337. doi: 10.1016/j.ygeno.2006.11.006

Kristjanson, PM, Swallow, Brent Murray, Rowlands, GJ, Kruska, RL, & De Leeuw, PN. (1999). Measuring the costs of African animal trypanosomosis, the potential benefits of control and returns to research. *Agricultural Systems, 59*(1), 79-98.

Kubatko, L. S., & Degnan, J. H. (2007). Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology, 56*(1), 17-24. doi: 10.1080/10635150601146041

Kulis, Marta, & Esteller, Manel. (2010). DNA methylation and cancer. *Advanced Genetics, 70*, 27-56.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174.

Law, J. A., & Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics, 11*(3), 204-220. doi: 10.1038/nrg2719

Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J. M., & Hoffmann, J. A. (1996). The dorsoventral regulatory gene cassette spatzle/Toll/cactus controls the potent antifungal response in Drosophila adults. *Cell, 86*(6), 973-983.

Lemey, P., Salemi, M., & Vandamme, A.M. (2009). *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*: Cambridge University Press.

Li, F. J., Gasser, R. B., Lai, D. H., Claes, F., Zhu, X. Q., & Lun, Z. R. (2007). PCR approach for the detection of *Trypanosoma brucei* and *T. equiperdum* and their differentiation from *T. evansi* based on maxicircle kinetoplast DNA. *Molecular and Cellular Probes, 21*(1), 1-7. doi: 10.1016/j.mcp.2006.03.009

Li, Long-Cheng, & Dahiya, Rajvir. (2002). MethPrimer: designing primers for methylation PCRs. *Bioinformatics, 18*(11), 1427-1431.

Lin, B. Z., Sasazaki, S., & Mannen, H. (2010). Genetic diversity and structure in Bos taurus and Bos indicus populations analyzed by SNP markers. *Animal Science Journal, 81*(3), 281-289. doi: 10.1111/j.1740-0929.2010.00744.x

Lizundia, R, Newman, C, Buesching, C.D, Ngugi, D, Blake, D, Sin, W.Y, . . . McKeeve, D. (2011). Evidence for a role of the host-specific flea (Paraceras melis) in the transmission of *Trypanosoma (Megatrypanum) pestanai* to the Europian Badger. *PLos ONE, 6*(2), e16977.

Loftus, R. T., MacHugh, D. E., Bradley, D. G., Sharp, P. M., & Cunningham, P. (1994). Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences of the United States of America, 91*(7), 2757-2761.

Lund, G., Andersson, L., Lauria, M., Lindholm, M., Fraga, M. F., Villar-Garea, A., . . . Zaina, S. (2004). DNA methylation polymorphisms precede any histological sign of atherosclerosis in mice lacking apolipoprotein E. *The Journal of Biological Chemistry, 279*(28), 29147-29154. doi: 10.1074/jbc.M403618200

Macdonald, D. W, & Newman, C. (2002). Population dynamics of badgers (Meles meles) in Oxfordshire, U.K.: numbers, density and cohort life histories, and a possible role of climate change in population growth. *Journal of Zoology, 256*, 121-138.

Macdonald, D. W., Anwar, M., Newman, C., Woodroffe, R., & Johnson, P. J. (1999). Inter-annual differences in the age-related prevalences of Babesia and *Trypanosoma* parasites of European badgers (Meles meles). *Journal of Zoology, 247*, 65-70. doi: DOI 10.1111/j.1469-7998.1999.tb00193.x

Magnus, E., Vervoort, T., & Van Meirvenne, N. (1978). A card-agglutination test with stained trypanosomes (C.A.T.T.) for the serological diagnosis of *T. B. gambiense* trypanosomiasis. *Annales de la Societe Belge de Medecine Tropicale, 58*(3), 169-176.

Maia da Silva, F, Noyes, H, Campaner, M, Junqueira, ACV, Coura, JR, Añez, N, . . . Teixeira, Marta MG. (2004). Phylogeny, taxonomy and grouping of *Trypanosoma rangeli* isolates from man, triatomines and sylvatic mammals from widespread geographical origin based on SSU and ITS ribosomal sequences. *Parasitology, 129*(05), 549-561.

Majekodunmi, A. O., Fajinmi, A., Dongkum, C., Picozzi, K., MacLeod, E., Thrusfield, M. V., . . . Welburn, S. C. (2013). Social factors affecting seasonal variation in bovine trypanosomiasis on the Jos Plateau, Nigeria. *Parasites & vectors, 6*, 293. doi: 10.1186/1756-3305-6-293

Mallatt, J. M., Garey, J. R., & Shultz, J. W. (2004). Ecdysozoan phylogeny and Bayesian inference: first use of nearly complete 28S and 18S rRNA gene sequences to classify the arthropods and their kin. *Molecular Phylogenetics and Evolution, 31*(1), 178-191. doi: 10.1016/j.ympev.2003.07.013

Mansfield, J. M., & Paulnock, D. M. (2005). Regulation of innate and acquired immunity in African trypanosomiasis. *Parasite Immunology, 27*(10-11), 361-371. doi: 10.1111/j.1365-3024.2005.00791.x

Maslov, D. A., Lukes, J., Jirku, M., & Simpson, L. (1996). Phylogeny of trypanosomes as inferred from the small and large subunit rRNAs: implications for the evolution of parasitism in the trypanosomatid protozoa. *Molecular and Biochemical Parasitology, 75*(2), 197-205.

Maslov, D. A., & Simpson, L. (1995). Evolution of Parasitism in Kinetoplastid Protozoa. *Parasitology Today, 11*(1), 30-32. doi: Doi 10.1016/0169-4758(95)80106-5

McGuire, K., Jones, M., Werling, D., Williams, J. L., Glass, E. J., & Jann, O. (2006). Radiation hybrid mapping of all 10 characterized bovine Toll-like receptors. *Animal Genetics, 37*(1), 47-50. doi: 10.1111/j.1365-2052.2005.01364.x

Medzhitov, R., & Janeway, C. A., Jr. (2000). How does the immune system distinguish self from nonself? *Seminars in Immunology, 12*(3), 185-188; discussion 257-344. doi: 10.1006/smim.2000.0230

Menzies, M., & Ingham, A. (2006). Identification and expression of Toll-like receptors 1-10 in selected bovine and ovine tissues. *Veterinary Immunology and Immunopathology, 109*(1-2), 23-30. doi: 10.1016/j.vetimm.2005.06.014

Meyer, A., Todt, C., Mikkelsen, N. T., & Lieb, B. (2010). Fast evolving 18S rRNA sequences from Solenogastres (Mollusca) resist standard PCR amplification and give new insights into mollusk substitution rate heterogeneity. *BMC Evolutionary Biology, 10*, 70. doi: 10.1186/1471-2148-10-70

Mishra, B. B., Gundra, U. M., & Teale, J. M. (2009). Toll-like receptors in CNS parasitic infections. *Current Topics in Microbiology and Immunology, 336*, 83-104. doi: 10.1007/978-3-642-00549-7_5

Mockenhaupt, F. P., Cramer, J. P., Hamann, L., Stegemann, M. S., Eckert, J., Oh, N. R., . . . Schumann, R. R. (2006). Toll-like receptor (TLR) polymorphisms in African children: common TLR-4 variants predispose to severe malaria. *The Journal of Communicable Diseases, 38*(3), 230-245.

Morley, E. K., Williams, R. H., Hughes, J. M., Terry, R. S., Duncanson, P., Smith, J. E., & Hide, G. (2005a). Significant familial differences in the frequency of abortion and *Toxoplasma gondii* infection within a flock of Charollais sheep. *Parasitology, 131*(Pt 2), 181-185.

Morley, E. K., Williams, R. H., Hughes, J. M., Terry, R. S., Duncanson, P., Smith, J. E., & Hide, G. (2005b). Significant familial differences in the frequency of abortion and *Toxoplasma gondii* infection within a flock of Charollais sheep. *Parasitology, 131*, 181-185. doi: Doi 10.1017/S0031182005007614

Morley, E. K., Williams, R. H., Hughes, J. M., Thomasson, D., Terry, R. S., Duncanson, P., . . . Hide, G. (2008). Evidence that primary infection of Charollais sheep with *Toxoplasma gondii* may not prevent foetal infection and abortion in subsequent lambings. *Parasitology, 135*(2), 169-173. doi: 10.1017/s0031182007003721

Mucha, R., Bhide, M. R., Chakurkar, E. B., Novak, M., & Mikula, I., Sr. (2009). Toll-like receptors TLR1, TLR2 and TLR4 gene mutations and natural resistance to Mycobacterium avium subsp. paratuberculosis infection in cattle. *Veterinary Immunology and Immunopathology, 128*(4), 381-388. doi: 10.1016/j.vetimm.2008.12.007

Muhanguzi, D., Picozzi, K., Hattendorf, J., Thrusfield, M., Kabasa, J., Waiswa, C., & Welburn, S. (2014). The burden and spatial distribution of bovine African trypanosomes in small holder crop-livestock production systems in Tororo District, south-eastern Uganda. *Parasites & Vectors, 7*(1), 603. doi: 10.1186/s13071-014-0603-6

Murphy, W. J., Eizirik, E., Johnson, W. E., Zhang, Y. P., Ryder, O. A., & O'Brien, S. J. (2001). Molecular phylogenetics and the origins of placental mammals. *Nature, 409*(6820), 614-618. doi: 10.1038/35054550

Murray, M., Morrison, W. I., & Whitelaw, D. D. (1982). Host susceptibility to African trypanosomiasis: trypanotolerance. *Advances in Parasitology, 21*, 1-68.

Naessens, J., Teale, A. J., & Sileghem, M. (2002). Identification of mechanisms of natural resistance to African trypanosomiasis in cattle. *Veterinary Immunology and Immunopathology, 87*(3-4), 187-194.

Nigerian Institute for Trypanosomiasis Research. (2008). Relevance to national development: Trypanosomiasis. Retrieved 23 March, 2015, from http://www.nitr-ng.org/relevant.html

Nimpaye, H., Njiokou, F., Njine, T., Njitchouang, G. R., Cuny, G., Herder, S., . . . Simo, G. (2011). *Trypanosoma vivax*, *T. congolense* "forest type" and *T. simiae*: prevalence in domestic animals of sleeping sickness foci of Cameroon. *Parasite, 18*(2), 171-179.

Njiokou, F., Nimpaye, H., Simo, G., Njitchouang, G. R., Asonganyi, T., Cuny, G., & Herder, S. (2010). Domestic animals as potential reservoir hosts of *Trypanosoma brucei gambiense* in sleeping sickness foci in Cameroon. *Parasite, 17*(1), 61-66.

Njiru, Z. K., Makumi, J. N., Okoth, S., Ndungu, J. M., & Gibson, W. C. (2004). Identification of trypanosomes in Glossina pallidipes and G. longipennis in Kenya. *Infection, Genetics and Evolution, 4*(1), 29-35. doi: 10.1016/j.meegid.2003.11.004

Nmorsi, Onyebiguwa Patrick Goddey, Isaac, Clement, Igbinosa, Igho Benjamin, Umukoro, Duncan Ogheneocovo, & Aitaikuru, Dafe Palmer. (2010). Human African trypanosomiasis in endemic focus of Abraka, Nigeria. *Asian Pacific Journal of Tropical Medicine, 3*(6), 448-450.

Noyes, HA, Ambrose, P, Barker, F, Begon, M, Bennet, M, Bown, KJ, & Kemp, SJ. (2002). Host specificity of *Trypanosoma (Herpetosoma)* species: evidence that bank voles (Clethrionomys glareolus) carry only one *T.(H.) evotomys* 18S rRNA genotype but wood mice (Apodemus sylvaticus) carry at least two polyphyletic parasites. *Parasitology, 124*(02), 185-190.

Ogunsanmi, AO, Ikeda, BO, & Akpavie, SO. (2000). Effects of management, season, vegetation zone and breed on the prevalence of bovine trypanosomiasis in southwestern Nigeria. *Israel Journal of Veterinary Medicine, 55*(2), 69-73.

Omotainse, S O, Edeghere, H, Omoogun, G A, Elhasan, E.O, Thompson, G A, Igweh, A C, . . . Halid, I. (2000). The prevalence of animal trypanosomosis in Konshisha local government area of Benue state, Nigeria. . *Israel Journal of Veterinary Medicine, 55*(4), 142-144.

Paling, R. W., Moloo, S. K., Scott, J. R., Gettinby, G., McOdimba, F. A., & Murray, M. (1991). Susceptibility of N'Dama and Boran cattle to sequential challenges with tsetse-transmitted clones of *Trypanosoma congolense*. *Parasite Immunology, 13*(4), 427-445.

Pariset, L., Mariotti, M., Marchitelli, C., Crisa, A., Napolitano, F., Failla, S., . . . Valentini, A. (2007). Polymorphisms within the Toll-like receptor (TLR)-2,-4, and-6 genes in cattle. *Italian Journal of Animal Science, 6*, 182-182.

Pearce, M A, & Neal, C. (1973). *Trypanosoma (Megatrypanum)* pestanai in British badgers (Meles meles). *International Journal of Parasitology, 4*(4), 439-440.

Peig, Jordi, & Green, Andy J. (2009). New perspectives for estimating body condition from mass/length data: the scaled mass index as an alternative method. *Oikos, 118*(12), 1883-1891. doi: 10.1111/j.1600-0706.2009.17643.x

Phillips, Theresa. (2008). The role of methylation in gene expression. *Nature Education, 1*(1), 116.

Picozzi, K., Carrington, M., & Welburn, S. C. (2008). A multiplex PCR that discriminates between *Trypanosoma brucei brucei* and zoonotic *T. b. rhodesiense*. *Experimental Parasitology, 118*(1), 41-46. doi: 10.1016/j.exppara.2007.05.014

Picozzi, K., Fevre, E. M., Odiit, M., Carrington, M., Eisler, M. C., Maudlin, I., & Welburn, S. C. (2005). Sleeping sickness in Uganda: a thin line between two fatal diseases. *BMJ (Clinical research ed.), 331*(7527), 1238-1241. doi: 10.1136/bmj.331.7527.1238

Playfair, J.H.L., & Bancroft, G.J. (2004). *Infection and Immunity*: Oxford University Press.

Poltorak, A., He, X., Smirnova, I., Liu, M. Y., Van Huffel, C., Du, X., . . . Beutler, B. (1998). Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science, 282*(5396), 2085-2088.

Pope, L. C, Butlin., R. K, Wilson, G. J, Woodroffe, R, Erven, K, Conyers, C.M, . . . Burke, T. (2007). Genetic evidence that culling increases badger movement: implications for the spread of bovine tuberculosis. *Molecular Ecology, 16*(23), 4919-4928.

Pujol-Rique, M., Derouin, F., Garcia-Quintanilla, A., Valls, M. E., Miro, J. M., & Jimenez de Anta, M. T. (1999). Design of a one-tube hemi-nested PCR for detection of *Toxoplasma gondii* and comparison of three DNA purification methods. *Journal of Medical Microbiology, 48*(9), 857-862.

Reiling, N., Holscher, C., Fehrenbach, A., Kroger, S., Kirschning, C. J., Goyert, S., & Ehlers, S. (2002). Cutting edge: Toll-like receptor (TLR)2- and TLR4-mediated pathogen recognition in resistance to airborne infection with Mycobacterium tuberculosis. *Journal of Immunology, 169*(7), 3480-3484.

Richard, G. F., Kerrest, A., & Dujon, B. (2008). Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiology and Molecular Biology Reviews 72*(4), 686-727. doi: 10.1128/mmbr.00011-08

Rogers, L. M., Delahay, R., Cheeseman, C. L., Langton, S., Smith, G. C., & Clifton-Hadley, R. S. (1998). Movement of badgers (*Meles meles*) in a high-density population: individual, population and disease effects. *Proceedings of the Royal Society B: Biological Sciences, 265*(1403), 1269-1276. doi: 10.1098/rspb.1998.0429

Salichos, L., & Rokas, A. (2013). Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature, 497*(7449), 327-331. doi: 10.1038/nature12130

Sandhu, G. S., Kline, B. C., Stockman, L., & Roberts, G. D. (1995). Molecular probes for diagnosis of fungal infections. *Journal of Clinical Microbiology, 33*(11), 2913-2919.

Sato, H., Al-Adhami, B. H., Une, Y., & Kamiya, H. (2007). *Trypanosoma (Herpetosoma) kuseli* sp. n. (Protozoa: Kinetoplastida) in Siberian flying squirrels (Pteromys volans). *Parasitol Research, 101*(2), 453-461. doi: 10.1007/s00436-007-0504-0

Seabury, C. M., Cargill, E. J., & Womack, J. E. (2007). Sequence variability and protein domain architectures for bovine Toll-like receptors 1, 5, and 10. *Genomics, 90*(4), 502-515. doi: 10.1016/j.ygeno.2007.07.001

Sengupta, P. P., Balumahendiran, M., Suryanaryana, V. V., Raghavendra, A. G., Shome, B. R., Gajendragad, M. R., & Prabhudas, K. (2010). PCR-based diagnosis of surra-targeting VSG gene: experimental studies in small laboratory rodents and buffalo. *Veterinary Parasitology, 171*(1-2), 22-31. doi: 10.1016/j.vetpar.2010.03.011

Shen, Lanlan, & Waterland, Robert A. (2007). Methods of DNA methylation analysis. *Current Opinion in Clinical Nutrition & Metabolic Care, 10*(5), 576-581.

Shiflett, A. M., Faulkner, S. D., Cotlin, L. F., Widener, J., Stephens, N., & Hajduk, S. L. (2007). African trypanosomes: intracellular trafficking of host defense molecules. *The Journal of Eukaryotic Microbiology, 54*(1), 18-21. doi: 10.1111/j.1550-7408.2006.00228.x

Simarro, Pere P, Diarra, Abdoulaye, Postigo, Jose A Ruiz, Franco, José R, & Jannin, Jean G. (2011). The human African trypanosomiasis control and surveillance programme of the World Health Organization 2000–2009: the way forward. *PLoS Neglected Tropical Diseases, 5*(2), e1007.

Soltis, P. S., Soltis, D. E., & Chase, M. W. (1999). Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature, 402*(6760), 402-404. doi: 10.1038/46528

Stefan, M., Zhang, W., Concepcion, E., Yi, Z., & Tomer, Y. (2014). DNA methylation profiles in type 1 diabetes twins point to strong epigenetic effects on etiology. *Journal of Autoimmunity, 50*, 33-37. doi: 10.1016/j.jaut.2013.10.001

Stevens, J. R., Teixeira, M. M., Bingle, L. E., & Gibson, W. C. (1999). The taxonomic position and evolutionary relationships of *Trypanosoma rangeli. International Journal for Parasitology, 29*(5), 749-757.

Steverding, Dietmar. (2008). The history of African trypanosomiasis. *Parasites & Vectors, 1*(3), 411.

Su, J., Wang, Y., Xing, X., Liu, J., & Zhang, Y. (2014). Genome-wide analysis of DNA methylation in bovine placentas. *BMC Genomics, 15*, 12. doi: 10.1186/1471-2164-15-12

Suzuki, M. M., & Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics, 9*(6), 465-476. doi: 10.1038/nrg2341

Takeet, M. I., Fagbemi, B. O., De Donato, M., Yakubu, A., Rodulfo, H. E., Peters, S. O., . . . Imumorin, I. G. (2013). Molecular survey of pathogenic trypanosomes in naturally infected Nigerian cattle. *Research in Veterinary Science, 94*(3), 555-561. doi: 10.1016/j.rvsc.2012.10.018

Tamura, K, Stecher, G, Peterson, D, Filipski, A, & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution, 30*, 2725-2729.

Taqi, M. M., Bazov, I., Watanabe, H., Sheedy, D., Harper, C., Alkass, K., . . . Bakalkin, G. (2011). Prodynorphin CpG-SNPs associated with alcohol dependence: elevated methylation in the brain of human alcoholics. *Addiction Biology, 16*(3), 499-509. doi: 10.1111/j.1369-1600.2011.00323.x

Taylor, T. K., Boyle, D. B., & Bingham, J. (2008). Development of a TaqMan PCR assay for the detection of *Trypanosoma evansi,* the agent of surra. *Veterinary Parasitology, 153*(3-4), 255-264. doi: 10.1016/j.vetpar.2008.01.045

Terry, R. S., Smith, J. E., Duncanson, P., & Hide, G. (2001). MGE-PCR: a novel approach to the analysis of *Toxoplasma gondii* strain differentiation using mobile genetic elements. *International Journal for Parasitology, 31*(2), 155-161.

Thumbi, S. M., McOdimba, F. A., Mosi, R. O., & Jung'a, J. O. (2008). Comparative evaluation of three PCR base diagnostic assays for the detection of pathogenic trypanosomes in cattle blood. *Parasites & Vectors, 1*(1), 46. doi: 10.1186/1756-3305-1-46

Torres-Garcia, D., Cruz-Lagunas, A., Garcia-Sancho Figueroa, M. C., Fernandez-Plata, R., Baez-Saldana, R., Mendoza-Milla, C., . . . Zuniga, J. (2013). Variants in toll-like receptor 9 gene influence susceptibility to tuberculosis in a Mexican population. *Journal of Translational Medicine, 11*, 220. doi: 10.1186/1479-5876-11-220

Truc, P., Jamonneau, V., N'Guessan, P., N'Dri, L., Diallo, P. B., & Cuny, G. (1998). *Trypanosoma brucei ssp*. and *T congolense*: mixed human infection in Cote d'Ivoire. *ransactions of the Royal Society of Tropical Medicine and Hygiene, 92*(5), 537-538.

Valarcher, J. F., & Taylor, G. (2007). Bovine respiratory syncytial virus infection. *Veterinary Research, 38*(2), 153-180. doi: 10.1051/vetres:2006053

Vanhamme, L., & Pays, E. (2004). The trypanosome lytic factor of human serum and the molecular basis of sleeping sickness. *International Journal for Parasitology, 34*(8), 887-898. doi: 10.1016/j.ijpara.2004.04.008

Vanhollebeke, B., Truc, P., Poelvoorde, P., Pays, A., Joshi, P. P., Katti, R., . . . Pays, E. (2006). Human *Trypanosoma evansi* infection linked to a lack of apolipoprotein L-I. *The New England Journal of Medicine, 355*(26), 2752-2756. doi: 10.1056/NEJMoa063265

Verma, Archana, Manchanda, Samiksha, Kumar, Nirmal, Sharma, Archna, Goel, Masha, Banerjee, Partha Sarathi, . . . Lejon, Veerle. (2011). *Trypanosoma lewisi* or *T. lewisi-like* infection in a 37-day-old Indian infant. *The American Journal of Tropical Medicine and Hygiene, 85*(2), 221-224.

Viola, L. B., Almeida, R. S., Ferreira, R. C., Campaner, M., Takata, C. S., Rodrigues, A. C., . . . Teixeira, M. M. (2009). Evolutionary history of trypanosomes from South American caiman (Caiman yacare) and African crocodiles inferred by phylogenetic analyses using SSU rDNA and gGAPDH genes. *Parasitology, 136*(1), 55-65. doi: 10.1017/s003118200800512x

Wastling, S. L., & Welburn, S. C. (2011). Diagnosis of human sleeping sickness: sense and sensitivity. *Trends in Parasitology, 27*(9), 394-402. doi: 10.1016/j.pt.2011.04.005

Welburn, S. C., Picozzi, K., Fevre, E. M., Coleman, P. G., Odiit, M., Carrington, M., & Maudlin, I. (2001). Identification of human-infective trypanosomes in animal reservoir of sleeping sickness in Uganda by means of serum-resistance-associated (SRA) gene. *Lancet, 358*(9298), 2017-2019.

Wiens, J J. (2006). Missing data and the design of phylogenetic analyses. *Journal of Biomedical Informatics. 39: 34–42., 39*, 34-42.

Williams, R. H., Morley, E. K., Hughes, J. M., Duncanson, P., Terry, R. S., Smith, J. E., & Hide, G. (2005). High levels of congenital transmission of *Toxoplasma gondii* in longitudinal and cross-sectional studies on sheep farms provides evidence of vertical transmission in ovine hosts. *Parasitology, 130*(Pt 3), 301-307.

World Health Organization. (2014). Trypanosomiasis, human African (sleeping sickness). Retrieved 23 March, 2015, from http://www.who.int/mediacentre/factsheets/fs259/en/

Wujcicka, W., Wilczynski, J., & Nowakowska, D. (2013). SNPs in toll-like receptor (TLR) genes as new genetic alterations associated with congenital toxoplasmosis? *European Journal of Clinical Microbiology & Infectious diseases, 32*(4), 503-511. doi: 10.1007/s10096-012-1763-y

Yang, Y., Li, X., Cui, W., Guan, L., Shen, F., Xu, J., . . . Gao, L. (2013). Potential association of pulmonary tuberculosis with genetic polymorphisms of toll-like receptor 9 and interferon-gamma in a Chinese population. *BMC Infectious Disiseases, 13*, 511. doi: 10.1186/1471-2334-13-511

Zhang, Z, Schwartz, S, Wagner, L, & Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *Journal of Computing Biology, 7*(1-2), 203-214.